

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 16:43:35 ; Search time 9219 Seconds  
(without alignments)  
11326.702 Million cell updates/sec

Title: US-09-915-706a-1

Perfect score: 3588

Sequence: 1 gcgcactatcgtatgatg.....gatnaattgcgcgttaa 3588

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	82.8	2.3	190050	1	AL646080	AL646080 Ralstonia
2	80	2.2	33154	1	AF361470	AF361470 Rhizobium
3	79.4	2.2	14867	3	AE001398	AE001398 Plasmid
4	76.2	2.1	179902	2	AP003630	AP003630 Oryza sat
5	75.4	2.1	143409	9	AL162497	AL162497 Human DNA
6	73	2.0	83703	5	AL732635	AL732635 Zebrafish
7	72.8	2.0	11422	6	AX345121	AX345121 Sequence
8	72.8	2.0	11422	6	AX348323	AX348323 Sequence
9	72.8	2.0	22419	2	AC119799	AC119799 Pan trogl
10	72.8	2.0	349980	6	AX344564	AX344564 Sequence
11	72.6	2.0	5727	1	AF037441	AF037441 Edwardsie
12	71.4	2.0	61020	6	AX251546	AX251546 Sequence
13	71.4	2.0	131033	5	AL591370	AL591370 Zebrafish
14	70.8	2.0	13568	1	AE009363	AE009363 Agrobacte
15	70.8	2.0	19961	1	AE008251	AE008251 Agrobacte
16	70.6	2.0	10029	1	AE013970	AE013970 Yersinia
17	70.6	2.0	72243	9	AL731858	AL731858 Human DNA
18	70.6	2.0	201050	9	AJ414143	AJ414143 Yersinia
19	70.2	2.0	101034	9	AC096550	AC096550 Homo sapi
20	69.8	1.9	86827	3	PFMA13P5	AL034556 Plasmid
21	69.4	1.9	8093	6	AX251190	AX251190 Sequence
22	69.4	1.9	8093	6	AX281167	AX281167 Sequence
23	69.4	1.9	8093	6	AX347411	AX347411 Sequence
24	69.4	1.9	8093	6	AX349132	AX349132 Sequence
25	69.4	1.9	8093	6	AX356412	AX356412 Sequence
26	69.4	1.9	131682	9	AL672277	AL672277 Human DNA
27	69.2	1.9	1141	6	AX083744	AX083744 Sequence
28	69	1.9	5504	6	AX252009	AX252009 Sequence
29	69	1.9	5504	6	AX348782	AX348782 Sequence
30	69	1.9	105574	2	AC013308	AC013308 Homo sapi
31	69	1.9	125097	9	AL691415	AL691415 Human DNA
32	69	1.9	161442	2	AL683875	AL683875 Homo sapi
33	69	1.9	162377	9	AL683870	AL683870 Homo sapi
34	68.8	1.9	191722	2	AC116360	AC116360 Homo sapi
35	68.6	1.9	837	6	A85629	A85629 Sequence 28
36	68.6	1.9	837	6	AR155122	AR155122 Sequence
37	68.6	1.9	837	6	E65647	E65647 Genome DNA
38	68.4	1.9	840	8	CNS0180K	AL110675 Botrytis
39	68.2	1.9	51014	2	AC011856	AC011856 Homo sapi
40	68.2	1.9	83391	6	AX458578	AX458578 Sequence
41	68	1.9	6306	6	AX278061	AX278061 Sequence
42	68	1.9	6306	6	AX323844	AX323844 Sequence
43	68	1.9	13131	6	AX356424	AX356424 Sequence
44	67.2	1.9	60604	2	AC023466	AC023466 Homo sapi
45	66.8	1.9	99003	2	AL390756	AL390756 Homo sapi

#### ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AL646080	AL646080	190050 bp DNA	AL646080	AL646080	segment 5/11.	Ralstonia solanacearum	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;	1 (bases 1 to 190050)	Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Arlat,M., Billault,A., Brotier,P., Camus,J.C., Cattolico,L.,

Pred. No. is the number of results predicted by chance to have a

TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
Chandler, M., Choise, N., Claudel-Renard, C., Cunnac, S., Demange, N., Gaspin, C., Lavie, M., Moisan, A., Robert, C., Saurin, M., Schlex, T., Sigulier, P., Thebaud, P., Whalen, M., Winker, P., Levy, M., Weissensbach, J. and Boucher, C.A.		Genome sequence of the plant pathogen <i>Ralstonia solanacearum</i> unpublished		2 (bases 1 to 190050)			
Boucher, C.A.		Direct Submission		Submitted (03-DEC-2001)	Genoscope and CNRS UMR-8030, 2 rue Gaston		
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie		Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,		Bp27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean			
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGW CNRS		118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA		URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,			
Laboratoire de Biometrie et Intelligence Artificielle INRA, Bp27,		F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire		INRA, Bp21, F31326 Castanet-Tolosan Cedex			
Christian.Bouchere@toulouse.inra.fr		http://sequence.toulouse.inra.fr/R.solanacearum.html.		Location/Qualifiers			
1. 190050							
/organism="Ralstonia solanacearum"							
/strain="GMT1000"							
/db_xref="taxon:305"							
/plasmid="megaplasmid"							
51. 362							
/gene="Rsp0643"							
/note="RS05579"							
51. 362							
/gene="Rsp0643"							
/note="Product confidence : hypothetical							
Gene name confidence : hypothetical							
Gene name confidence : hypothetical							
predicted by Codon-usage							
predicted by Framed"							
/codon_start=1							
/evidence=not_experimental							
/transl_table=1							
/product="HYPOTHETICAL PROTEIN"							
/protein_id="CAD17794.1"							
/db_xref="GI:17431114"							
/translation="MKTSPENPAAGPASPDEAAGDGOVAFOELHEYRNHTITRPR							
RIGGGLMRAPAAVSLDSRALVAVTEVGQARPTIGAAADAVDAARSPIDARLACARP							
A"							
397. 555							
/gene="Rsp0644"							
/note="RS05578"							
397. 555							
/gene="Rsp0644"							
/function="miscellaneous; unknown"							
/note="Product confidence : putative							
Gene name confidence : hypothetical							
predicted by Codon-usage							
predicted by Framed"							
/codon_start=1							
/evidence=not_experimental							
/transl_table=1							
/product="PUTATIVE SIGNAL PEPTIDE PROTEIN"							
/protein_id="CAD17795.1"							
/db_xref="GI:17431115"							
/translation="MERQQLFAIVIVIAVSAASLARDLNDDEDRPPQGAALVHGHH							
ATASAAR"							
complement(575. 1135)							
/gene="Rsp0645"							
/note="RS05577"							
complement(575. 1135)							
/gene="Rsp0645"							
/function="miscellaneous; not classified regulator"							
/note="Product confidence : probable							
Gene name confidence : hypothetical							
predicted by Codon-usage							

predicted by Homology  
 predicted by "framed"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="PROBABLE TRANSCRIPTION REGULATOR PROTEIN"  
 /protein\_id="CAD17796.1"  
 /db\_xref="GI:17431116"  
 /translation="MKRISTPEPVDKKVIAIKFRSPAAPGQASALDFTDRALIKMLQDASISYNALADKYSLPSPACKLRVEVLEKLGIRGVALLERKALDGLTVMI  
 GVLDRSPESPADKEKAVOKVPGCLCEHVVTGSEDFYMLVTRTNDNSYNRLHAQELL  
 YLPGVROIRTFMTLKVQVLSITQLPIA"  
 1244..2260  
 /gene="RSP0646"  
 /note="RS05576"  
 1244..2260  
 /gene="RSP0646"  
 /EC\_number="4.1.99.4"  
 /function="miscellaneous; hypothetical/global homology"  
 /note="Product confidence: probable  
 Gene name confidence: hypothetical  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by "framed"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="PROBABLE 1-AMINOCYCLOPROPANE-1-CARBOXYLATE  
 DEAMINASE (ACC DEAMINASE) PROTEIN"  
 /protein\_id="CAD17797.1"  
 /db\_xref="GI:17431117"  
 /translation="MNLKHPRHPLTGTGPPPIQPLKRLSAHLGGKVELYAKREDCSNG  
 LAFGNKTRKLELYEVLVAGGYDTVTSIGISQNRQYAAVAHLGICVYQEWV  
 VYAAVAVDRNGINELSLTGADVRILAAGPDIGIRSMRQAMDVVRAGGCEPIPA  
 GCSERPLGGLGFGVGAEEVROGEAFEGFRPPYIYCVSTGSGTQGMVYGAAGGRADR  
 VTIGDASAKPEQTRQQLIRIARDTRAKLVEIKRGLTDEDDVLDTRTGGPEYGLNEGTL  
 EAIRIKAQEGELTDPRVEKSGMHGMDIRVNGGEFPGESRYLVANHLGGVPAALNAYSFL  
 FRNG"  
 complement(2336..3508)  
 /gene="RSP0647"  
 /note="RS05575"  
 complement(2336..3508)  
 /gene="RSP0647"  
 /EC\_number="4.2.1.17"  
 /function="miscellaneous; hypothetical/global homology"  
 /note="product confidence: putative  
 Gene name confidence: hypothetical  
 predicted by Codon\_usage  
 predicted by Homology  
 predicted by "framed"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /transl\_table=11  
 /product="PUTATIVE ENOYL-COENZYME A HYDRATASE PROTEIN"  
 /protein\_id="CAD17798.1"  
 /db\_xref="GI:17431118"  
 /translation="MSALPKPAEPAAAEPEVLPHVYVNRVAIITLNRPSALNLSHQM  
 VROLSALIAOCRRDPRITAAVYLRGAGEKFGAGGDVRAIYLRAGOGAHPDGMLOFF  
 VDERLIDYLAHREKRPVYALLDGTAMGGGMLGCGAALRYVTESSRIAMPETRTIGLP  
 DVGAIHFGAMPPEMALITGLTGTTLSCADLHCGGLADVCVPAWMLASREDRLRLISH  
 AGDDPDLQALRAVREPCCNTVPAHALARTPTMVARFDVRSSTIDRIYVALRSLSDH  
 PAREKRLVQKTLDMATHSPMLHVTREALRLRQTLLECFEMELGIVARAIEEED  
 FCEGVRHLVQTDKDRGMAPATLAGVREVRVHFLSSPWSGAAMPLAHLGAQDQA"  
 complement(3505..4305)  
 /gene="RSP0648"  
 /note="RS05574"  
 complement(3505..4305)  
 /gene="RSP0648"  
 /EC\_number="4.2.1.17"  
 /function="small molecule metabolism; degradation; carbon  
 compounds"  
 /note="Product confidence: probable  
 Gene name confidence: hypothetical

predicted by Codon\_usage  
predicted by Homology  
predicted by Framed  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE ENOYL-COA HYDRATASE PROTEIN"  
/protein\_id="CAD17799.1"  
/db\_xref="GI:17431119"  
/translation="MIALDWLHDTGVALTLISRPANAFTEPREGIROLQATVEITLADP  
RVRALVITGDSKPKFSACADLNTASGRDVARERAAFGAFELQANRPVITAIN  
GYAMGGELCALACDRIAEHRAQLAVEITAVGLCCGGTQLPWLVEGWAKMIL  
TGERIDAQTLALIGLEVEVDIGARREALAMARVITLSRAVTFSKALIHQARHG  
PRTAALAVERRERFVDFGHPDREGVNAFLERPRPMHTAARPEELP"  
complement(4302..5198)  
/gene="nmsB"  
/note="RSP0649; RS05573"  
complement(4302..5198)  
/gene="nmsB"  
/EC\_number="1.1.1.31"  
/function="small molecule metabolism; amino acid  
biosynthesis; isoleucine/valine"  
/note="Product confidence : probable  
Gene name confidence : putative  
predicted by Codon\_usage  
predicted by Homology  
predicted by Framed  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="PROBABLE 3-HYDROXYISOBUTYRATE DEHYDROGENASE  
OXIDOREDUCTASE PROTEIN"  
/protein\_id="CAD17800.1"  
/db\_xref="GI:17431120"  
/translation="MKIAFGLGNKGMARNLKAGHALSVFEDLDARAVOSLVADAGA  
SAABPKAAATGADCAITMLPAAHVSILVADGILAGIAKVPILIDSSITDPDSAK  
ALALATAHGNPFDVAPVSGGTGAAGCTLFPMGGAAFEQVAPVLSAMGKNLVHC  
GEAGGOGAKICNNILVIGITMAGVAEAMSLGEALIDIKRVGIIINTSGRMSDIT  
NPFGVETAPSSRGTGGFGFDLMLKDLGLAGDAKAKVRQVYLGALAOOLYQASSS  
OGAKIDFSAVIRLYRPAAGAR"  
complement(5220..6767)  
gene  
Query Match 2.3%; Score 82.8; DB 1; Length 190050;  
Best Local Similarity 56.3%; Pred. No. 3.9e-07;  
Matches 175; Conservative 0; Mismatches 133; Indels 3; Gaps 1;  
gene  
3275 CGTGCAGCATCACTATGACGTTGAACGAATGCCGCCGTAAAGACGAAGACTGCCG 3334  
11  
DB 112111 CGCGTGACACTGACGACGCTTGAAGGGCGACGATCGACGAGAACTGCGG 112170  
11  
QY 3335 TTGTGTTGGGGTCATTGGGCACTTTTCAGACACAAACCAAGATCAGAAAAAGTTGAT 3394  
11  
DB 112171 TTGCGTGGTGGGCTGGCGGACCTGCGGCCCACT--CGGAAGTGGCGCAGCCCAAG 112227  
11  
QY 3395 TTAGAAGACGAGAGTTCACGGGTATCGATAAAGCAACTTCATGATGGGCGAA 3454  
11  
DB 112228 CCGCGGACCGGAGAGTTCCTCAACATCGACCGCAACTTCGACGACCTGATGCGCG 112287  
11  
QY 3455 ATTACCCCGCGTCTTCTGACAAAGTTGATTAACAAGCTTGCTAATGATGATAGCCAGTTT 3514  
11  
DB 112288 GCCGAGCCCGCGCTGCTTTAGATGCGAGAACACGCTGACGGAAGAGGCGCGCTTC 112347  
11  
QY 3515 GAAGTGAAGTTCGCTCGGTGATGAAGATTTCCACCGCAGAGAACTTACTGTATNAA 3574  
11  
DB 112348 GCGCGTGCACCTCAAGTTCCGCTGCTGAGGATTCTCGCCGGAAGCCGTTGTAACAG 112407  
11  
QY 3575 ATTGAGCCGCT 3585  
11  
DB 112408 ATCGAGCCGCT 112418  
11  
RESULT 2  
AF361470

LOCUS AF361470 33154 bp DNA linear BCT 23-OCT-2001  
DEFINITION Rhizobium leguminosarum bv. trifolii imp gene locus, complete  
SEQUENCE AF361470  
ACCESSION AF361470  
KEYWORDS AF361470.1 GI:16326458  
SOURCE  
ORGANISM Rhizobium leguminosarum bv. trifolii.  
Rhizobium leguminosarum bv. trifolii.  
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
Rhizobiaceae: Rhizobium  
1 (bases 11567 to 12038)  
Roest,H.P., Mulders,I.H., Spalink,H.P., Wijffelman,C.A. and  
Lugtenberg,B.J.  
A Rhizobium leguminosarum biovar trifolii locus not localized on  
the sym plasmid hinders effective nodulation on plants of the pea  
cross-inoculation group  
Mol. Plant Microbe Interact. 10 (7), 938-941 (1997)  
97449849  
MEDLINE 9304865  
JOURNAL  
REFERENCE 2  
AUTHORS 2 (bases 1 to 33154)  
Bladergroen,M.R., Badelt,K., Stronk,O.P., Lugtenberg,E.J.J. and  
Spalink,H.P.  
An avirulence locus of a symbiotic Rhizobium leguminosarum strain  
is involved in temperature-dependent protein secretion  
Unpublished  
3 (bases 1 to 33154)  
Bladergroen,M.R., Badelt,K., Stronk,O.P., Lugtenberg,E.J.J. and  
Spalink,H.P.  
Direct Submission  
Submitted (15-MAR-2001) Institute of Molecular Plant Sciences,  
Leiden University, Wassenaarseweg 64, Leiden 2333 AL, The  
Netherlands  
FEATURES  
source  
gene  
CDS  
location/Qualifiers  
1..33154  
/organism="Rhizobium leguminosarum bv. trifolii"  
/db\_xref="taxon:386"  
/note="biovar: trifolii"  
744..1820  
/gene="impa"  
744..1820  
/gene="impa"  
/function="Involved in temperature-dependent protein  
secretion"  
/note="hypothetical"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="ImpA"  
/protein\_id="MAL17808.1"  
/db\_xref="GI:16326486"  
/translation="MNVRREITDPLEQNIHPCGNNRNSATFEITYRIRIDARNARTAE  
RSIIPEGITTTAPAWHDVSNLGLDLSKSDIEVLAWLAERQLRQFSGLHDVYA  
TVSLDKHFDALHSIGDGVVERPAPFGLNGVGEGLTIGAIRLTSILPGKFAQFS  
LMDFLDSORPNETERRKELQOAAEAGVAGOMSAVLDVYTCGIAAFDRMVELIDECGD  
OAPPSNTRNVLOEASAIIRMIAGIEATEPAPLAPOLASOGSEPLPASFEQGEAI  
GLRPSAMIRISREAEFELLAVARYFRTRPSPHSISITIVLRGMDFELLAEI  
LPBQGTAAVLTAGIOPVADKGC"  
1857..2402  
/gene="impB"  
1857..2402  
/gene="impB"  
/function="Involved in temperature-dependent protein  
secretion"  
/note="hypothetical; similar to Edwardsiella ictaluri  
hypothetical protein Eip20"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=11  
/product="ImpB"  
/protein\_id="MAL17797.1"  
/db\_xref="GI:16326475"  
/translation="MGRRIISMASVHERKLERKPRVHIKYEVEETEGAAVVKELPFV  
GLGDFSGNPTQPLKPFGERKRVQVDRDNFDVVMRMPGLTLSEVNTLEKNGTELPA

SLKESPMADPGGVVDOVGLKALLANRQJLBDLMSKRD8SEELERLEIDLQNKNA  
LSOLVAQGGKKADEADALN"  
2533..4014  
/gene="ImpC"  
2533..4014  
/gene="ImpC"  
/function="involved in temperature-dependent protein  
secretion"  
/note="hypothetical; similar to Edwardsiella ictaluri  
hypothetical protein Eip55"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="ImpC"  
/protein\_id="AA117801.1"  
/translation="MPAEQLOLRKEDVGFAPDENILSKVAATROTQEDPDRAOLLRTLL  
TDQALKGIVYDRRLRTLLNNAIAELDKAISQIAIMQSPETFLBAGMGILYIKK  
NSETSNALKITVNLASKRDVDAKLAAVERDOSRLFKSYDEBETPGEGPMALIGG  
LEPNDSKEDVQVLQGVSMIAAFAAPFISAASPMGFEDFEKLAKPDLKIEYVW  
YAKMSSLDSDSPSPVTLAMPRYLAMPYGPKPPIEEENPDPEGSGTGLSGSYC  
MMNATVYGTGTLTAFPSKNGMCTAIGAENGKVEGLPMHIFSSDGDLDLCKPT  
ITDRDEALGELGELPLCHIKNDYVFPFGAQRAPKLYDRPEATNAAVASARLPIW  
MATSFAHYIKVMGRDKISGEMADECEAMLNINISNYNANDDGEESRAKRYPLREAR  
KVYVQELPGRGATNVAAMRPWLQMBELTSLRMVARLPSKN"  
4081..5469  
/gene="ImpD"  
4081..5469  
/gene="ImpD"  
/function="involved in temperature-dependent protein  
secretion"  
/note="hypothetical; similar to ImpC and to Edwardsiella  
ictaluri hypothetical protein Eip55; possible duplication  
of ImpC"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="ImpD"  
/protein\_id="AA117802.1"  
/translation="MGSGSTPMRRARQDLITLIDEALNKOVNALEHHPQAMAEKRWG  
LAMIVRSRSGEYKVLGIVTSELTRMERASDPDQSHLPELILYSERCMGSGEPL  
CLIVGDTYLSPADAGDAPSTISQIGWAAAAPCFPVGAGAPLAIGLEDEHLELRNVA  
DFSWLANSTAIKRWGRAREDSRFLGLVAPRLMKSPJLTPGSSRNKNGPFFRSIAG  
DGSLLSKMAFAATVYINPFIIDSMFADIRGVQDAVAGGLSTAEPLPYDGTGSG  
DGLSKQAVEHLEHSGEQLCELGIIIPATYIATSSAIFENQSLAHPHYSEMKAR  
ONARLAAPLOVLCASPSHYIKVIMRDEGLOLAISIERKLAWLITAYLTGGDDG  
IALTRPRLBSAGICVAIEKRGKSGSCYVRQDPHQDLDVATSHLLAETATGGQILL  
SRPAVPERKSA"  
5466..6296  
/gene="ImpE"  
5466..6296  
/gene="ImpE"  
/function="involved in temperature-dependent protein  
secretion"  
/note="hypothetical"  
/codon\_start=1  
/evidence=not experimental  
/transl\_table=11  
/product="ImpE"  
/protein\_id="AA117803.1"  
/translation="MPTLSAISQSLSDNALDAIEEVKAHLKVRSPQEAHRLITDILL  
/transl="G1:16326481"  
VLAGVQAGNOCSTIATLSPDATMGFPALLRNPLRAMARQWMTSGALPFPQGS  
LDKLAVALRGIAGRGNNDKARTLDALEKLRGRPLINNGKAVSPFRDLDRTPTHALE  
VIMTGGGLWIDPKATIASLIEPIARPDIAERKALSLIDAAASVLLPVIYGTGK  
DETALRGHEWEIETTGITSGGRCYLAGEMWSFHDITQSLIEVPAAGAROVYAGG

[illegible]



QY	2074	ACTTCGACGAGCTCATGGCGCGCATGACGCCGGCGCTCACCTGTCGGGTGCAGAACTC	2133
QY	3492	TTGCTAATGATGATGATGACCGATTTGTAAGTGAACCTTGACCTCCGTTGCATGAAGAATTTCC	3551
Db	2134	TCGAGAGAAACGGAACAGACGCTTCCCGTTTGGCTGAATTTGCAAAAGCATGCCGATTTCC	2193
OY	3552	ACCGAGAACTTGATGATNAATTTAGCCGCGCTTAA	3588
Db	2194	AACCGGCGCGGCTGCTGCATGAGTGCCTGCCGGCGTGA	2230
RESULT 3	AE001398		
LOCUS	AE001398	14867 bp	DNA linear INV 16-Apr-2002
DEFINITION	Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.		
ACCESSION	AE001398	AE001362	
VERSION	AE001398.1	GI:3845197	
KEYWORDS			
SOURCE	Plasmodium falciparum 3D7.		
ORGANISM	Plasmodium falciparum 3D7		
REFERENCE	Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
AUTHORS	1 (bases 1 to 14867) Gardner M.J., Tettelin H., Cerucci D.J., Cummings L.M., Aravind L., Koonin E.V., Shalimov S., Mason T., Yu K., Fujii C., Pederson J., Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Bereta M., Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O., Fraser C.M., Adams M.D., Venter J.C. and Hoffman S.L.		
TITLE	Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum		
JOURNAL	Science 282 (5391), 1126-1132 (1998)		
MEDLINE	95021743		
PUBMED	9804551		
REFERENCE	2 (bases 1 to 14867) Gardner M.J.		
AUTHORS	Direct Submission		
TITLE	Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1..14867		
SOURCE	/organism="Plasmodium falciparum 3D7"		
	/strain="3D7"		
	/db_xref="taxon:36329"		
	/chromosome="2"		
	complement(1570..2424)		
	/gene="PFB0490c"		
	complement(1570..2424)		
	/gene="PFB0490c"		
	/note="predicted by GlimmerM"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="AAC71887.1"		
	/db_xref="GI:3845198"		
	/translation="MKKKNKINDYLSCEPDVDYDRKKSGKSNLSKSSSTKSKSDYRK SLSFKRDSHKKGSSFRGRKSGTINKSSSEFKPYTNRLINNTYNNVGRNHNHR DNKGRGRSGFSGVFDNRKSEFKRFLSNKSSVKSYSRNGSKMGKSKFNKAPTSSR TWTYKRLNNYKTYTAPYKPKFNNNLISLYRRNRFALTNRKSKPYGCTKSSVPRRIK GLKSKSLKSTKRSKTSKSGSKFKPLNKYFLSKIKITVSLNKLPSLKEQKNTSEVNLPESL NNNTTKKN"		
	11241..14606		
	/gene="PFB0495w"		
	11241..14606		
	/gene="PFB0495w"		
	/note="predicted by GlimmerM"		
	/codon_start=1		
	/product="hypothetical protein"		
	/protein_id="AAC71888.1"		
	/db_xref="GI:3845199"		
	/translation="MFALKRNTVREGFVNICFSLYKLILYKSNFVTVNLVETNNRKR TKNKIKSKAOSLFEDGLNIHDLILFKLIPKIKCAKYECISAEYKYLIDYKKGC ENYISLIDTIOSYKIFDELDKTFTDVFYEVKNIDKNVCKINIEYFKKKDTPHRR ETIICKLTIOSYKIHENKNGDLHFLIFPRWKNKDNLLFTYTYTNYVFDHMYLEN HEIYKLFETFNKTLNNNSNIPFNKNLIQEMFNLTYRETIKNEKNYIILKMKKEIYRK		

[illegible]



repeat\_region /note="LIMB9 repeat: matches 5789. .6085 of consensus"  
2061. .2106 /note="23 copies 2 mer tg 91% conserved"  
repeat\_region 2063. .2106 /note="11 copies 4 mer tggt 93% conserved"  
repeat\_region 2138. .2422 /note="LIMB3 repeat: matches 6691. .7023 of consensus"  
repeat\_region 2422. .3051 /note="LIMC4 repeat: matches 7168. .7837 of consensus"  
repeat\_region 3674. .4545 /note="LIPB4 repeat: matches 5272. .6144 of consensus"  
repeat\_region 5701. .5910 /note="7 copies 30 mer 89% conserved"  
repeat\_region 6490. .6613 /note="MIR repeat: matches 92. .225 of consensus"  
repeat\_region 9599. .9648 /note="25 copies 2 mer aa 94% conserved"  
repeat\_region 10251. .10559 /note="ALUSX repeat: matches 2. .312 of consensus"  
repeat\_region 11134. .11521 /note="LIMB7 repeat: matches 5762. .6161 of consensus"  
repeat\_region 12667. .12710 /note="11 copies 4 mer tggt 81% conserved"  
repeat\_region 13210. .13309 /note="L2 repeat: matches 2559. .2674 of consensus"  
repeat\_region 13794. .14148 /note="THELC repeat: matches 1. .371 of consensus"  
repeat\_region 14816. .15030 /note="MIR repeat: matches 49. .256 of consensus"  
repeat\_region 14982. .15037 /note="L2 repeat: matches 2651. .2705 of consensus"  
repeat\_region 15476. .15940 /note="L2 repeat: matches 1314. .1835 of consensus"  
repeat\_region 16305. .16609 /note="ALUSX repeat: matches 1. .300 of consensus"  
repeat\_region 16683. .16722 /note="20 copies 2 mer tt 77% conserved"  
repeat\_region 16948. .17073 /note="MLTIC repeat: matches 348. .464 of consensus"  
repeat\_region 17074. .17360 /note="ALUB repeat: matches 13. .299 of consensus"  
repeat\_region 17361. .17661 /note="MLTIC repeat: matches 7. .348 of consensus"  
repeat\_region 17974. .18347 /note="LIMD1 repeat: matches 5659. .6029 of consensus"  
repeat\_region 18399. .18554 /note="3 copies 52 mer 75% conserved"  
repeat\_region 18400. .18549 /note="5 copies 30 mer 73% conserved"  
repeat\_region 18413. .18552 /note="5 copies 28 mer 74% conserved"  
repeat\_region 18565. .18738 /note="LIMB2 repeat: matches 6150. .6335 of consensus"  
repeat\_region 19070. .19105 /note="9 copies 4 mer gata 88% conserved"  
repeat\_region 19363. .19664 /note="L2 repeat: matches 2436. .2750 of consensus"  
repeat\_region 20032. .20053 /note="11 copies 2 mer aa 100% conserved"  
repeat\_region 21272. .21354 /note="ALUSQ repeat: matches 1. .283 of consensus"  
repeat\_region 22754. .22918 /note="3 copies 55 mer 77% conserved"  
repeat\_region 26601. .26684 /note="L2 repeat: matches 2668. .2748 of consensus"  
repeat\_region 26685. .26814 /note="ALUB repeat: matches 2. .131 of consensus"  
repeat\_region 26815. .27110 /note="ALUSQ repeat: matches 1. .303 of consensus"  
repeat\_region 27111. .27285 /note="ALUB repeat: matches 130. .307 of consensus"  
repeat\_region 27286. .27578 /note="L2 repeat: matches 2359. .2668 of consensus"

repeat\_region .29945. .30257 /note="ALUY repeat: matches 1. .306 of consensus"  
repeat\_region 30812. .31078 /note="ALUSX repeat: matches 1. .267 of consensus"  
repeat\_region 31887. .32091 /note="MER58A repeat: matches 1. .222 of consensus"  
repeat\_region 32818. .32853 /note="9 copies 4 mer aac 100% conserved"  
repeat\_region 32856. .32957 /note="MER97C repeat: matches 185. .289 of consensus"  
repeat\_region 32958. .33406 /note="MLTIC repeat: matches 1. .466 of consensus"  
repeat\_region 33407. .33867 /note="MER97C repeat: matches 289. .726 of consensus"  
repeat\_region 33895. .34718 /note="8 copies 103 mer 61% conserved"  
repeat\_region 33915. .34519 /note="121 copies 5 mer atata 56% conserved"  
repeat\_region 33920. .34103 /note="92 copies 2 mer at 57% conserved"  
repeat\_region 33924. .34553 /note="21 copies 30 mer 61% conserved"  
repeat\_region 33925. .34730 /note="13 copies 62 mer 61% conserved"  
repeat\_region 33940. .34709 /note="14 copies 55 mer 61% conserved"  
repeat\_region 33954. .34213 /note="13 copies 20 mer 70% conserved"  
repeat\_region 33999. .34698 /note="25 copies 28 mer 58% conserved"  
repeat\_region 34014. .34650 /note="13 copies 49 mer 62% conserved"  
repeat\_region 34071. .34642 /note="11 copies 52 mer 62% conserved"  
repeat\_region 34300. .34709 /note="205 copies 2 mer aa 56% conserved"  
repeat\_region 34390. .34689 /note="15 copies 20 mer 60% conserved"  
repeat\_region 34488. .34689 /note="53 copies 4 mer taata 59% conserved"  
repeat\_region 34512. .34703 /note="64 copies 3 mer ata 60% conserved"  
repeat\_region 34629. .34718 /note="3 copies 30 mer 78% conserved"  
repeat\_region 35062. .35647 /note="LIMB1 repeat: matches 5550. .6144 of consensus"  
repeat\_region 35694. .35885 /note="L2 repeat: matches 2054. .2250 of consensus"  
repeat\_region 35882. .36229 /note="L2 repeat: matches 1446. .1832 of consensus"  
repeat\_region 36377. .36484 /note="MIR repeat: matches 48. .147 of consensus"  
repeat\_region 39838. .39877 /note="20 copies 2 mer tt 82% conserved"  
repeat\_region 39841. .39875 /note="7 copies 5 mer tggtt 85% conserved"  
repeat\_region 41466. .41514 /note="L2 repeat: matches 2659. .2747 of consensus"  
repeat\_region 41792. .42143 /note="THEIB repeat: matches 1. .364 of consensus"  
repeat\_region 43471. .43590 /note="L2 repeat: matches 2629. .2750 of consensus"  
repeat\_region 44937. .45157 /note="L2 repeat: matches 2259. .2489 of consensus"  
repeat\_region 45491. .45781

Query Match 2.1%; Score 75.4; DB 9; Length 143409;  
Best Local Similarity 47.6%; Pred. No. 1.4e-05;  
Matches 313; Conservative 0; Mismatches 341; Indels 3; Gaps 3;  
DY 461 ATTGCGTACATATATACTTTGCTATCTGAAAGCTTAATGTCAGGAGCGGCTTGGT 520  
DB 34685 ATTTATATTAAATATATATTTATATATATATTTATATATATATATATATATATATTA 34626



ORGANISM	synthetic construct
REFERENCE	artificial sequences.
AUTHORS	Olek, A., Piepenbrock, C. and Berlin, K.
TITLE	Diagnosis of diseases associated with the immune system
JOURNAL	Patent: WO 0200928-A 192 03-JAN-2002;
FEATURES	Epigenomics AG (DE)
SOURCE	Location/Qualifiers
	1. .11422
	/organism="synthetic construct"
	/db_xref="taxon:32630"
	/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT	3913 a 59 c 1856 g 5554 t
ORIGIN	
Query Match	2.0%; Score 72.8; DB 6; Length 11422;
Best Local Similarity	47.7%; Pred. No. 5.4e-05;
Matches 244; Conservative	0; Mismatches 267; Indels 1; Gaps 1;
Oy	605 AATTATCATATATATATGATGCTATTTATCTCTGATATGATCATCTTATGCT 664
Db	9660 AATTATATTTTTTAAATATAGTGTATGCGTTATAGTATATAGATTAATATGCTAGT 9719
Oy	665 CTATCATGTAATATATTTTCGAGTACCTTAATTCAGCAATTTGCTATTAATA 724
Db	9720 TAAATGCTTATATATATTTTTTTTGTATTTTATTTTATGATTTAAATTTTATTA 9779
Oy	725 TTATCTGAATAATATGTAATGATGCTTTGTGTTATTTATGTTTGTTCATTTT 784
Db	9780 TATTTTGAATTATGCTGATATATATATATATGAGTATATATGTTATGTTATTTT 9839
Oy	785 ATGACGGTGAGCTTGTCATTCATATTTTATATGATGACACATCTTTGATGAGTATTT 844
Db	9840 GTTATTTATATGATGATATATAATAATTTTGTAGTAATAATATATTTGAACTATTTT 9899
Oy	845 AAGATATTTGTAATATGATCAGAGGCGTTGCGTATTTT-TTATATTAATCATATTA 903
Db	9900 TATTTATTTTGCATATATATGAATTTATTTAAATTTATGATTTATTTTAAATTA 9959
Oy	904 TCACAAATATATGTTATTTTGTCTTTTATATAGTGTCTTTTAAAGAGGTAGATGACC 963
Db	9960 TTATTAATTTTATTTTATATTTTGAACAATAATAGTTATTAAGAGGTTAAGTATTTGTTT 10019
Oy	964 TAAAGTGCGCTTAATATATGCGCTAAATTTGCCATTCCTATATATTCACCTCA 1023
Db	10020 TAAATTAATATAGATATTCAGAGTTACTTTTGTATTTTATATATATTAATAAGTA 10079
Oy	1024 ATTGCAAAATTGACAAATATGTCACCTTGTGTAATGAACAATATATAGTGTGTTTG 1083
Db	10080 TTTTGTATATGCAATATTTGAGAGTATTTTAAATTTTAAATTTTAAATTTTAA 10139
Oy	1084 CTGCAAAATATAAAATTTTTCGTGTTGAATA 1115
Db	10140 TTGGAATATATTAATACGATATGATTAAGA 10171
RESULT 8	
LOCUS	AX348323 11422 bp DNA linear PAT 06-FEB-2002
DEFINITION	Accession 18 from Patent WO0202806.
ACCESSION	AX348323
VERSION	AX348323.1 GI:18614359
KEYWORDS	
ORGANISM	synthetic construct.
SOURCE	artificial sequences.
REFERENCE	1
AUTHORS	Olek, A., Piepenbrock, C. and Berlin, K.
TITLE	Method and nucleic acids for pharmacogenomic methylation analysis
JOURNAL	Patent: WO 0202806-A 18 10-JAN-2002;
FEATURES	Epigenomics AG (DE)
SOURCE	Location/Qualifiers
	1. .11422

BASE COUNT	3913 a	59 c	1856 g	5594 t								
ORIGIN	/organism="synthetic construct" /db_xref="taxon:32630" /note="Chemically treated genomic DNA (Homo sapiens)"											
Query Match	2.0%	Score 72.8:	DB 6:	Length 11422:								
Best Local Similarity	47.7%	Pred. No. 5.4e-05:										
Matches 244:	Conservative	0:	Mismatches 267:	Indels 1:	Gaps 1:							
OY	605	AATTTCACATATATATGATGCTATTAATTAATCTGTCGTATGGTATCAACCTTCATGCT	664									
DB	9660	AATTATATTTTAAATATATGCTATGCTGATATAGTATATAGATAAATATATGCTAGT	9719									
OY	665	CTATACATGTAATATATTTGAGTAGACCTTAATTCAGGTAAATTTCTATTTAATTA	724									
DB	9720	TAAATGGTTATATATATTTTGTTTTATTTTATTTATATTAATTTAATTTAATAA	9779									
OY	725	TTATCTGAATATATATGATATGATGCTTTGGGTATTTTATGTTCTTTCATTTTA	784									
DB	9780	TATTTTGAATTAATGATATATATATTAATATAGGTAATTAATGATTTATTTATTTT	9839									
OY	785	ATGACGGAGACCTTGATCATATTTTATTTATGATGACACATCTTGATGATGATTT	844									
DB	9840	GTTATTTATATGATCATATATATATTTTATGATATTAATATATTTTGAAGTATATTT	9899									
OY	845	AAGATATTTGTAATGATCATGAGGGCTTGCGTATTTT-TTATATTTAATCATATAA	903									
DB	9900	TATTTATTTTGAATATATGAATTTTATTAATTTATGATATTTTAAATTAAGTATTA	9959									
OY	904	TCAACAATATATGTTATTTTGTCTTTTATATGCTCTTTTAAAGAGTATGATGAC	963									
DB	9960	TTATTAATTTTATTTTATATTTTGAAGAAATATAGTATTAAGAGTTTAAATTTGTT	10019									
OY	964	TAAAGTGCGCCATTAATGAGCGTAAATTTGCCCTTCTTAATTCACCTCAAGATAC	1023									
DB	10020	TAAATTTATATAGATATTCAGGTTACTTTTGTATTTTATATGATTTATTAATAG	10079									
OY	1024	ATTGGCAATTTGACAAATATATGCTACTGCTATGAACAATATTTAGTATGTTGTT	1083									
DB	10080	TTTGTATATATGATATTTGAGTGATTTTATTAATTTTATTTTATTTTATTTT	10139									
OY	1084	CTGCAAAATTAATAATTTTTCGTGTTGAATA	1115									
DB	10140	TTGGAATTAATTAATATACGATATGATTAAGTA	10171									
RESULT 9												
AC119799/c												
LOCUS	AC119799	224149 bp	DNA	linear	HTG 21-AUG-2002							
DEFINITION	Pan troglodytes clone ch251-9K16, WORKING DRAFT SEQUENCE, 9											
ACCESSION	AC119799											
VERSION	AC119799.10	GI:22380735										
KEYWORDS	HTG; HTGS; PHASE1; HTGS-DRAFT.											
SOURCE	chimpanzee.											
ORGANISM	Pan troglodytes											
REFERENCE												
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;											
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.											
REFERENCE												
AUTHORS	Gao,N., Hu,X., Eichler,E. and Roe,B.A.											
TITLE	Pan troglodytes BAC Clone ch251-9K16											
JOURNAL	Unpublished											
REFERENCE												
AUTHORS	2 (bases 1 to 224149)											
TITLE	Gao,N., Hu,X., Eichler,E. and Roe,B.A.											
JOURNAL	Direct Submission											
REFERENCE												
AUTHORS	Submitted (02-MAY-2002) Department Of Chemistry and Biochemistry,											
TITLE	The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,											
JOURNAL	OK 73019, USA											
REFERENCE												
AUTHORS	3 (bases 1 to 224149)											
TITLE	Gao,N., Hu,X., Eichler,E. and Roe,B.A.											
JOURNAL	Direct Submission											
REFERENCE												
AUTHORS	Submitted (21-AUG-2002) Department Of Chemistry and Biochemistry,											
TITLE												
JOURNAL												

The University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA  
On Aug 21, 2002 this sequence version replaced gl:21541857.

----- Genome Center  
Center: Department of Chemistry And Biochemistry  
The University of Oklahoma  
Center code:OKMOR

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 9 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 4434: contig of 4434 bp in length  
\* 4435 4534: gap of unknown length  
\* 4535 8606: contig of 4072 bp in length  
\* 8607 8706: gap of unknown length  
\* 8707 20579: contig of 11873 bp in length  
\* 20580 20679: gap of unknown length  
\* 20680 39415: contig of 18736 bp in length  
\* 39416 39515: gap of unknown length  
\* 39516 66441: contig of 26926 bp in length  
\* 66442 66541: gap of unknown length  
\* 66542 96035: contig of 29494 bp in length  
\* 96036 96135: gap of unknown length  
\* 96136 126068: contig of 29933 bp in length  
\* 126069 126168: gap of unknown length  
\* 126169 169210: contig of 43042 bp in length  
\* 169211 169310: gap of unknown length  
\* 169311 224149: contig of 54839 bp in length.

## FEATURES

Location/Qualifiers

1..224149  
/organism="Pan troglodytes"  
/db\_xref="taxon:9598"  
/clone\_id="CHORI-251 Chimpanzee BAC library"

BASE COUNT 61562 a 46989 c 47682 g 67109 t 807 others  
ORIGIN

Query Match 2.0%; Score 72.8; DB 2; Length 224149;  
Best Local Similarity 50.8%; Pred. No. 5.1e-05;  
Matches 199; Conservative 0; Mismatches 192; Indels 1; Gaps 1;

```
OY 523 TTTTATGCTAGCGCATATTCATTAATTCCTAATAATAAGCTCTAAAGC 582
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 39946 TATATATGATATATCATCTATATATATATATATGATATATCATATATTAATATACAC 39887

OY 583 TGTATATGATGATATTCGAAATTAACATATATATGCTATTAATTAATCTGCTG 642
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 39886 TACTATATATATATATATATATATATATATATATATATATATATATATATATATG 39887

OY 643 TATTCGATACACTTCGCTACATGATATATATTCGATGACCTTAATCA 702
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 39826 TATTTATATATATATATATATATATATATATATATATATATATATATATATATTTT 39767

OY 703 AGTAATATGCTATATATATATATATGATATATGATATATGCTCTGCTGCTAT 762
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 39766 TTAATTTTATATATATATATATATATATATATATATATATATATATATATATTTT 39707

OY 763 TTTTATGCTTTCATTTTATATACAGCTGCTGCTATATATTTTATATGAGA 822
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 39706 TTAATTTTATATATATATATATATATATATATATATATATATATATATATATATAT 39647

OY 823 CAACATCTTGATGAGATAT-7AAGATATGTTATGATGAGGCGTTGCGGTATTT 881
    | | | | | | | | | | | | | | | | | | | | | | | |
Db 39646 TATATATATATATATATATATATATATATATATATATATATATATATATATATAT 39587

OY 882 TTTATATTAATCAATTAATAATCAACAATAT 913
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 39586 ACCATATATTAATATATATATATATATAT 39555
```

## RESULT 10

AX344564

LOCUS AX344564 349980 bp DNA linear PAT 01-FEB-2002  
DEFINITION Sequence 15 from Patent WO0200932.

AX344564

ACCESSION AX344564.1 GI:18492450  
VERSION  
KEYWORDS

SOURCE

ORGANISM

synthetic construct.  
artificial sequences.

REFERENCE

AUTHORS

Olek, A., Piepenbrock, C. and Berlin, K.  
Diagnosis of known genetic parameters within the mhc  
Patent: WO 0200932-A 15 03-JAN-2002;

JOURNAL

Epidemiol

Genetics AG (DE)

FEATURES

source

Location/Qualifiers

1..349980

/organism="synthetic construct"

/db\_xref="taxon:32630"

/note="chemically treated genomic DNA (Homo sapiens)-original length of seq 1: 3,673,778 &lt;223&gt;-split as follows.-seq 01 0.000.001 TO 0.349.980-seq 02 0.300.001 649.980-seq 03 600.001 949.980-seq 04 900.001 1.249.980-seq 05 1.200.001 1.549.980-seq 06 1.500.001 1.849.980-seq 07 1.800.001 2.149.980-seq 08 2.100.001 2.449.980-seq 09 2.400.001 2.749.980-seq 10 2.700.001 3.049.980-seq 11 3.000.001 3.349.980-seq 12 3.300.001 3.649.980-seq 13 3.600.001 3.673.778 &lt;223&gt;-original length of seq 2: 3.673,778 &lt;223&gt;-split as follows.-seq 14 0.000.001 TO 0.349.980-seq 15 0.300.001 649.980-seq 16 600.001 949.980-seq 17 900.001 1.249.980-seq 18 1.200.001 1.549.980-seq 19 1.500.001 1.849.980-seq 20 1.800.001 2.149.980-seq 21 2.100.001 2.449.980-seq 22 2.400.001 2.749.980-seq 23 2.700.001 3.049.980-seq 24 3.000.001 3.349.980-seq 25 3.300.001 3.649.980-seq 26 3.600.001 3.673.778"

BASE COUNT 99616 a 3868 c 72947 g 173548 t 1 others  
ORIGIN

Query Match 2.0%; Score 72.8; DB 6; Length 349980;  
Best Local Similarity 50.0%; Pred. No. 5e-05;  
Matches 235; Conservative 0; Mismatches 232; Indels 3; Gaps 2;

```
OY 455 CGATTAATTCGCTGATATACCTTTCATCGAAGCTTAATGCTGAGGACGCG 514
    ||||| | | | | | | | | | | | | | | | | | | | | | | |
Db 264188 CGATTAATTCATTCGCTTATATTTTATATGTTAATGATGATTAATAT 264247

OY 515 TTTGCTGCTTAATTCGTTACGCTCATATTCATTAATTCATTAATTAACAGTTG 574
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 264248 TTTTATGCTTAATTCGTTATTTTATATATTTTATGGAATATTTTAAAGTTT 264307

OY 575 TAAAGCGCTTTATGATGATATTCGAATATATCATATATATATGATGATATTT 634
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 264308 TGTTCGTTTAAATTCGATTTTAAATGCTTTTTCGAGTTTGAAGTTT 264367

OY 635 ACTTCGCTATTCGATACCTTCATGCTCATATGATATATATTCAGTTAGAC 694
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 264368 TATATTTAGATATAGTTTTCGTTATATATATATATATATATATATATATAT 264427

OY 695 TTAATTCAT-AGTAATATGCTATATATATATATATATATATATATATATATAT 752
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 264428 GTAGTTTATTTTATATATATATATATATATATATATATATATATATATATATAT 264487

OY 753 TTTCGCTATTTTATATGCTTTTCATTTTAAAGCGCTGACCTTGCAATCATATTT 812
    | | | | | | | | | | | | | | | | | | | | | | | |
Db 264488 GAAATTTATATTTATAGTTTTTTTATATATATATATATATATATATATATATAT 264547

OY 813 TTTATGATGACAACATCTTCATGATGATATTAAGATATGTTATATGATGAGGCTTTG 872
    || | | | | | | | | | | | | | | | | | | | | | | |
Db 264548 TTTTCGTTTATGATTTATAT-TTTAATATATTTTATATTTTAAATATTAAGGTTT 264606
```









repeat_region	6196. 6443	/note="Pr0000036 repeat: matches 23. .271 of consensus"
repeat_region	6449. .6507	/note="Pr00000341 repeat: matches 703. .764 of consensus"
repeat_region	6507. .6579	/note="Pr00000149 repeat: matches 323. .395 of consensus"
repeat_region	6584. .6870	/note="Pr0000149 repeat: matches 1. .284 of consensus"
repeat_region	6669. .7150	/note="Pr0000183 repeat: matches 39. .315 of consensus"
repeat_region	7154. 7371	/note="dir459 repeat: matches 1. .218 of consensus"
repeat_region	7369. 7489	/note="dir814 repeat: matches 6. .126 of consensus"
repeat_region	7497. .7611	/note="dir813 repeat: matches 3. .119 of consensus"
repeat_region	8146. 8200	/note="Pr0000306 repeat: matches 290. .353 of consensus"
repeat_region	8200. .8326	/note="Pr0000105 repeat: matches 130. .256 of consensus"
repeat_region	8517. .8774	/note="ANGEL repeat: matches 3. .301 of consensus"
repeat_region	8781. 9197	/note="Pr000298 repeat: matches 23. .442 of consensus"
repeat_region	9438. .9734	/note="Pr0000403 repeat: matches 1. .362 of consensus"
repeat_region	9631. .9746	/note="Pr0000380 repeat: matches 4. .118 of consensus"
repeat_region	9992. 9957	/note="5.8 copies 12 mer TTTATATATATTA 80% conserved"
repeat_region	10092. .10333	/note="Pr0000407 repeat: matches 28. .356 of consensus"
repeat_region	10968. .11078	/note="Pr0000365 repeat: matches 14. .127 of consensus"
repeat_region	11132. .11208	/note="Pr000403 repeat: matches 1. .76 of consensus"
repeat_region	11204. .11475	/note="Pr000396 repeat: matches 2. .301 of consensus"
repeat_region	12115. .12369	/note="Pr0000155 repeat: matches 1. .261 of consensus"
repeat_region	12370. .13175	/note="dir460 repeat: matches 1. .806 of consensus"
repeat_region	13178. .13236	/note="Pr000155 repeat: matches 258. .375 of consensus"
repeat_region	13485. .13607	/note="dir474 repeat: matches 7. .139 of consensus"
repeat_region	13666. .13713	/note="dir474 repeat: matches 317. .363 of consensus"
repeat_region	13714. .13934	/note="match: STS: Em:G39804"
repeat_region	13715. .14262	/note="Pr0000004 repeat: matches 1. .552 of consensus"
repeat_region	14169. .14297	/note="match: STS: Em:G41123"
repeat_region	14315. .14472	/note="Pr0000004 repeat: matches 456. .576 of consensus"
repeat_region	14489. .14667	/note="dir475 repeat: matches 151. .309 of consensus"
repeat_region	14666. .15479	/note="Pr0000183 repeat: matches 1. .178 of consensus"
repeat_region	15361. .15905	/note="Pr0000061 repeat: matches 1. .831 of consensus"
repeat_region		/note="CPG island"
repeat_region	15480. .16154	/note="Pr0000142 repeat: matches 1. .675 of consensus"
repeat_region	15480. .15604	/note="match: STS: Em:G41434"
repeat_region	16153. .16180	/note="Pr0000285 repeat: matches 542. .569 of consensus"
repeat_region	16178. .16428	/note="Pr0000183 repeat: matches 173. .424 of consensus"

```

repeat_region      16553. .16748
/noise="TDRI9 repeat: matches 95. .294 of consensus"
repeat_region      16743. .16816
/noise="ANGEL repeat: matches 123. .195 of consensus"
repeat_region      16893. .17036
/noise="drr466 repeat: matches 7. .148 of consensus"
repeat_region      17182. .17355
/noise="Dr000365 repeat: matches 1. .182 of consensus"
repeat_region      17508. .17812
/noise="Dr000183 repeat: matches 1. .315 of consensus"
misc_feature       complement(17586. .17719)
/noise="match: STRS: Em:G45723"
repeat_region      17721. .17869
/noise="Dr000341 repeat: matches 613. .763 of consensus"
repeat_region      18365. .18708

Query Match      2.0% Score 71.4; DB 5; Length 131033;
Best Local Similarity 49.1%; Pred. No. 0.0001;
Matches 189; Conservative 0; Mismatches 196; Indels 0; Gaps 0;
```

QY	612	ACATATAATATGATGACCTATATATCTGCTGATTTGGATATCAACATCTTCATGCTATAC	671
Db	92176	ACCTTAGATATTATAGAAATAGTTTTTTTTTTAGGATTTTATCATTTTATATTTATTTA	92235
QY	672	TGTATATATATTCGAGTTAGACCTAAATTCACAGTAATTTGCTATTTATATATTTATCG	731
Db	92236	TTTATATATATTTATTTATATTTATATTAATTTATTTATTTATTTAAATTTATTTATTTT	92295
QY	732	AATAATATATGATCATTCGTTTGCGTATTTTATGCTTTTTCATTTTAAATGACGG	791
Db	92296	ATTTTGTGTTAAATTTTTTATATTTTATTTATTTTATAGTACTTTTGTAAAAAATATTTAGT	92355
QY	792	TGAGCTTGTCATTCATATTTTATATGATGACACACTCTTGATGCAAGATTTAAGATAT	851
Db	92356	ATTTATTTTTTTGTTTTTTTTTTTTTTTTTATTTATTTATGTAATATTTTATTTTCATTTAT	92415
QY	852	TGTTATGATCATGAGGGTTTGCGTATTTTATTTATTAATCAATATAATTAATCAACAT	911
Db	92416	TAAATATTTTATATTTTCTATATTTTGTCTTATTTATTTTAAATATTTTAAAT	92475
QY	912	ATATGTTATTTTGTCTTTTATATAGTCTCTTTTAAAGAGTAGATGACCTAAAGTCT	971
Db	92476	TAAATTTACTTTTAAATATTTATTTTGTGTTTATTTACTATGTAATATTTATGATGATAC	92535
QY	972	GCCTAATATGCGCTAAATGGCAT	996
Db	92536	GTAATTAATATATTAATTTAACTT	92560

RESULT	14
AEO09363/c	
LOCUS	
DEFINITION	AEO09363 13568 bp DNA linear BCT 20-DEC-2001
ACCESSION	Agrobacterium tumefaciens str. C58 linear chromosome, section 133 of 187 of the complete sequence.
VERSION	AEO09363 AEO08689
KEYWORDS	AEO09363.1 GI:17742806
SOURCE	.
ORGANISM	Agrobacterium tumefaciens str. C58 (U. Washington).
REFERENCE	Agrobacterium tumefaciens str. C58 (U. Washington). Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium. 1 (bases 1 to 13568)
AUTHORS	Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E., Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F., Zhou,Y., Boyce Sr.,D., Chapman,P., Glendening,J., Deatherage,G., Gillet,M., Grant,C., Guenther,D., Kennedy,T., Levy,R., Li,M., McClelland,E., Palmieri,A., Raymond,C., Rouse,G., Seoplimmachak,C., Wu Z., Gordon,D., Eisen,J.A., Paulsen,I., Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M., Kravpan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C., Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.Y., and Nester,E.W.
TITLE	The genome of the natural genetic engineer Agrobacterium

JOURNAL	21608550	11743193	2 (bases 1 to 13568)	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.E., Zhou, Y., Boye, S.D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmeri, A., Raymond, C., Rouse, G., Saenplimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tinney, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.
REFERENCE	Submitted (27-SEP-2001)	Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA		
TITLE	Journal			
FEATURES	source	Location/Qualifiers		
gene	CDS	1. 13568 /organism="Agrobacterium tumefaciens str. C58 (U. Washington)" /strain="C58" /db_xref="taxon:180835" complement(149. .1585) /gene="Atu4340" complement(149. .1585) /gene="Atu4340" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAI45134.1" /db_xref="GI:17742807" /translation="MISHPGPIHRSPLMAHOSPPARIRADLIIDALITGQVNA ILIHPDQAEGRKGLAMVREAGSTDVKVLNLSWRDLAKNTERADFDQSHLE EYVNRFEFGMGEPVGLMIGDYAFSPDLDGASITTLQIGVANAACPFVAAA PSAGVLDSELSRVHDFSSLKQESRLRNALRARDSEFGVLAIRVLLRSPIRY ARGDGPFPREREHVAESGFTLLMGNAFAFATVYVNRVDSGMFADLRGVTOADVG MSLPRLPDDLTIESNGLSAOPPEVQVLTGCGOQSDGIYVSTTYSGAIRFA NOSTLHAGHSTEARONARLALMLOVLCASRSHLKYIMRDIGQLSDATITTEK LESWLSYTLGNDDAESILRTYPLRSAGINVEVPKGTFCSTYVLOPHQDLVDVS TSPHLIAETMMPPAAASRSASAPORISA" complement(1598. .3079) /gene="Atu4341" complement(1598. .3079) /gene="Atu4341" complement(1598. .3079) /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAI45135.1" /db_xref="GI:17742808" /translation="MSAESLKSSEAQTATNAENOGLSKVVASATROTEPPDRANLRLT, TDALKTGYVDNRLVYTLNHLAIELDRITSEDLAALMOAPEAKLEGTRAGNYLVK NSESTVNLKIRVMNAGKRELARDLEKAVEFDOSRLFKAIYEDEGTPGSPGLAIIID YERGNSEPDVOLLGVSSITAAAPFPTISASPHMGFEYRDLARDRDEKFEVYE YAKRSFRESDDSRFVTLALPRYLARPYGKTPNPVDDFAYDERGAVNDDLDHDEC WMAAAVYMGKTLTEAFKSGMCTAIKGAENGGRVENLPMHVFSDDGDLDLKCPTEG ITDRDAELKGLFGLCHYKNTDYAVFEGAOTAHKPKLYDKPEATANAASVLRPLV KATVVEVPGKPGVYNAVAMRPMLOMEELTSLRMVARISKNS" complement(3173. .3709) /gene="Atu4342" complement(3173. .3709) /gene="Atu4342" complement(3173. .3709) /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11		
gene	CDS	2. 13568 /organism="Agrobacterium tumefaciens str. C58 (U. Washington)" /strain="C58" /db_xref="taxon:180835" complement(149. .1585) /gene="Atu4340" complement(149. .1585) /gene="Atu4340" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAI45136.1" /db_xref="GI:17742809" /translation="MPRGIKVEVMAVSEHEKLEVRKRPRVHIKYEVEGAMVKEJPF VVGVLDFSGNPNPLKPFGEKFEVQIDRNPFDNPRMPLNIGTAVNTLQDGTDM QVNLKFSMDPEFGAVYGVPAKLLDLARNELROLLSKADSDELDRLDLQNK TDALVLSLSEKNGASE" complement(3748. .4803) /gene="Atu4343" complement(3748. .4803) /gene="Atu4343" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAI45137.1" /db_xref="GI:17742810" /translation="MDTQVHKRDIIEFGNCGNIRNDSRTREIYYIKDERNARAE RAASPDONLKISSWDSVSNLQIILYSESKDVEILLAMLEASLRGLRGEKREIEL CGDLFYNHMDLSKSIISDNDDEKFAFGAGINGISDEGTIOPRLASLIPGKFGESH LMPFOAORNEKSRRELYRIASEGAVMASSHLAIVNTCLSPFATVAVLSEKQ AAPSSNIRMTLIEAAAIRTLIGRDEPAPVBPOTPAIAGTDESGOSARTPASPE GISSRDEAFETLSVARYPRTRPHSPISISITLVIRGRNDESELLAELLPTQARN AVLTAGIRKPGGDNNGK" 5024. .7702 /gene="clpB" /note="Atu4344" 5024. .7702 /gene="clpB" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="ATP-dependent Clp protease, ATP-binding subunit" /protein_id="AAI45138.1" /db_xref="GI:17742811" /translation="MSHIDILNRLVGLAEPDLRVTLLEAASVAVRMGRHYVDIPWLLA VVAGIAYFEFDLKIPLPYLAKIEISLEEAIGGEALISQNTLTAREWMLAS LEAGRDVLCDDLLMDEETSLRSTYRSFSAFPLKAMDRCALRSSAGAGVYVP SALTSGSEASQAAGONDRLRTYTHDMTLDARNGVDPVYIGRDEDLRLVDLITRRR ONNPILVGEAGVKRTVAEALAEFASGNVPEKLRVCLINDISILQACAGKGEFE RRLHGYIDAVKRSAEVILFTIDAHGLVGAAGAGGDDANILKPAKRGVETVATV TWSEYKIFEKDALLRRQPVHVRDEDTAIRMGLVADTFVSHNNTVROEAIVA AVOLASRYMPAROLPDKAVALDITAAAVSLAROTPERLRAMESRHLISDELNML REPODEMONRQISIRDELERLFIAGIDDLGRYDAEMAEALAESEOPATGASVNSH LRPATREPMANARLIVPTVVDREALIAAVYRWGIGLKLAPOLISARTLDVMBOR VVQDDAIRTILADAMRTARGLSDPRPRPVPFLVYVCSGKRETNALSLADLLYGENS HLTTINKSEFKESHKAVSLIGSPPGVYVGRGGVLAIVARRFVGLDLDEIDKHPG VODIFVQVDFKGVLRDGEGRDVPKNTTFTMTANTGSELISLSDAPDVPRESEALD LAMPLELKQFPKPAFLKRTIILPMPPLGAELASVVMQIGLIDRVLATVGTDLRLSD AARDALVAVRAGASEIGARATEIMGIDLEPLSSFFLEKVIAGEVRGKIVVDGEENG F GIRAEPAAGEDEFAVTEVGVDAASDCAITRRMRH" 7760. 8236 /gene="Atu4345" 7760. 8236 /gene="Atu4345" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAI45139.1" /db_xref="GI:17742812" /translation="MPLIYLOIDIGODATHEHRKMDIEAIHMNVARNMNTSAGSAA NREASPTISEVILTVSDDSSSKRLQOENACAGTGLATILHTVTNPGNTYIEVLLT NTLIASVSDSSGDRPEVETIKLNFYMEKAYITPFDNNSPQSMIASYDLATTKAA" 8283. .8732 /gene="Atu4346" 8283. 8732 /gene="Atu4346" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer"		
gene	CDS	3. 13568 /organism="Agrobacterium tumefaciens str. C58 (U. Washington)" /strain="C58" /db_xref="taxon:180835" complement(149. .1585) /gene="Atu4340" complement(149. .1585) /gene="Atu4340" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAI45139.1" /db_xref="GI:17742813" /translation="MPLIYLOIDIGODATHEHRKMDIEAIHMNVARNMNTSAGSAA NREASPTISEVILTVSDDSSSKRLQOENACAGTGLATILHTVTNPGNTYIEVLLT NTLIASVSDSSGDRPEVETIKLNFYMEKAYITPFDNNSPQSMIASYDLATTKAA" 8283. .8732 /gene="Atu4346" 8283. 8732 /gene="Atu4346" /note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer"		





Mon Mar 3 10:44:09 2003

us-09-915-706a-1.rge

Page 17

Search completed: March 2, 2003, 02:21:44  
Job time : 15022 secs

---



GenCore version 5.1.4-p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 08:42:34 ; Search time 734 Seconds  
(without alignments)  
11008.406 Million cell updates/sec

Title: US-09-915-706a-1

Perfect score: 3588  
Sequence: 1 gtcgactatcgtcgtggtg.....gatnaattgagccgcttaa 3588

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: N.Geneseq\_101002.\*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*
- 25: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	72.8	2.0	11422	24	ABK39937 Human chemically p
2	72.8	2.0	11422	24	ABL32219 Human immune syste
3	71.4	2.0	61020	22	AAS46788 Tumour suppressor
4	69.4	1.9	8093	22	AAS46436 Tumour suppressor
5	69.4	1.9	8093	24	ABL9237 Chemically treated
6	69.4	1.9	8093	24	ABK3974 Human DNA for stag
7	69.4	1.9	8093	24	ABL49332 Human polynucleoti
8	69	1.9	5504	24	ABL70350 Chemically treated
9	69	1.9	5504	24	AAS61309 Human gene regulat

10	68.2	1.9	83391	24	ABO67094 Human angiogenesis
11	68	1.9	6306	22	AAS45516 Chemically pretrea
12	68	1.9	6306	24	ABK28458 DNA transcription
13	68	1.9	13131	24	ABL92349 Chemically treated
14	66.6	1.9	8666	22	AAS46305 Tumour suppressor
15	66.6	1.9	8666	24	ABO67177 Human anglogenesis
16	66.6	1.9	8666	24	ABK34008 Human DNA for stag
17	66.6	1.9	8666	24	ABL32396 Human immune syste
18	66.6	1.9	9504	24	ABK28407 DNA transcription
19	66.2	1.8	17848	22	AAS45323 Chemically pretrea
20	66.2	1.8	17848	24	ABK39976 Human chemically p
21	66.2	1.8	17848	24	ABK28164 DNA transcription
22	66.2	1.8	18624	24	ABL33702 Human immune syste
23	66	1.8	19124	18	AAT72882 Plasmodium var-7 g
24	66	1.8	19124	21	AAS298287 Plasmodium var-7 g
25	65.8	1.8	7849	24	ABL92279 Chemically treated
26	65.8	1.8	7849	24	AAD22330 Chemically treated
27	65.8	1.8	19459	24	ABL70527 Chemically treated
28	65.8	1.8	19459	24	ABK31212 Signal transductio
29	65.4	1.8	1179	24	ABO49604 Oligonucleotide fo
30	65.4	1.8	1179	24	ABO49605 Oligonucleotide fo
31	65.2	1.8	6040	24	ABL92300 Chemically treated
32	65.2	1.8	6040	24	ABL49369 Human polynucleoti
33	65	1.8	6048	24	ABO67002 Human anglogenesis
34	65	1.8	9539	22	AAS45347 Chemically pretrea
35	65	1.8	9539	24	ABK28180 DNA transcription
36	65	1.8	9642	24	ABL32357 Human immune syste
37	65	1.8	12237	24	ABL34358 Human immune syste
38	64.8	1.8	9905	24	ABL32062 Human immune syste
39	64.8	1.8	11555	24	ABL32616 Human immune syste
40	64.8	1.8	11555	24	AAD28380 Human chemically t
41	64.8	1.8	18183	22	AAS46279 Tumour suppressor
42	64.8	1.8	18183	24	ABL70111 Chemically treated
43	64.8	1.8	18183	24	ABK31158 Signal transductio
44	64.6	1.8	6233	22	AAS46497 Tumour suppressor
45	64.6	1.8	7571	24	ABL32527 Human immune syste

ALIGNMENTS

RESULT 1	ABK39937	standard; DNA: 11422 BP.
ID	ABK39937	
AC	ABK39937	
XX		
XX		
DT	21-MAY-2002 (first entry)	
XX		
DE	Human chemically pretreated gene sequence #9 strand 2.	
XX		
KW	Human: ds; bisulphite treatment; CPG; DNA methylation; cancer; tumour;	
KW	Cytostatic; ALDH6; CYP11A; CYP11B; CYP3A3; DPVD; EPHX2; OCLN; TXNRD1;	
KW	UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.	
OS	Homo sapiens.	
XX		
PN	WO200202806-A2.	
XX		
PD	10-JAN-2002.	
XX		
PF	29-JUN-2001; 2001WO-EP07470.	
XX		
PR	30-JUN-2000; 2000DE-1032529.	
XX		
PA	01-SEP-2000; 2000DE-1043826.	
XX		
PI	(EPIC-) EPIGENOMICS AG.	
XX		
PI	Olek A, Piepenbrock C, Berlin K;	
XX		
DR	WPI; 2002-154757/20.	
XX		
PT	New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,	

PT useful for detecting cytosine methylation state of genes associated  
PT with pharmacogenomics and for therapy of diseases e.g. cancer  
PS Claim 1; SEQ ID No 18; 24pp; English.

XX  
CC The invention relates to a nucleic acid comprising a sequence at  
CC least 18 bases in length of a segment of the chemically pretreated DNA  
CC of genes associated with pharmacogenomics according to one of the  
CC sequences of the genes ALDH6 (NM\_000693), CYP11A (NM\_000781), CYP11B1  
CC (NM\_000497), CYP3A3 (NM\_000776 and NM\_017460), DPYD (NM\_000110), EPHX2  
CC (NM\_001979), OCLN (NM\_002538), TXNRD1 (NM\_003330), UGT8 (NM\_003360),  
CC MRP (NM\_004996, NM\_019900, NM\_019901, NM\_019902, NM\_019862, NM\_019898,  
CC NM\_019899) and their complementary sequences, or a sequence (SI) chosen  
CC from 87 sequences and their complements. The chemical pretreatment  
CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)  
CC into uracils. Also included are an oligomer (II) in particular an  
CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in  
CC each case at least one base sequence having a length of 9 nucleotides  
CC which hybridises to or is identical to a chemically pretreated DNA of  
CC genes associated with pharmacogenomics and their complements, arranged in  
CC an array for analysing diseases associated with the methylation state  
CC (CPG) and/or detecting SNPs (single nucleotide polymorphisms)  
CC of the 87 sequences. The oligomers may also be used as PCR primers.  
CC The set of 87 nucleic acids and their complements is useful for diagnosis  
CC and therapy of solid tumours and cancer. The present sequence  
CC represents one the 87 DNA sequences or its complement.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pcl\_sequences.

XX  
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 other;

Query Match 2.0%; Score 72.8; DB 24; Length 11422;

Best Local Similarity 47.7%; Pred. No. 4.1e-07;

Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

OY 605 AATTATACATAAATATGATGCTATTATATGCTGATGCTATACACTTCATGCT 664  
DB 9660 AATTAAATTTTAAATGATGCTATGCTATGCTATATATAGATAATATAGTGTGT 9719  
OY 665 CTATACATGTAATATATGATGCTATGCTATGCTATGCTATGCTATGCTATGCT 724  
DB 9720 TAAATGCTTAT 9779  
OY 725 TTAATGCAAT 784  
DB 9780 TATTTTAT 9839  
OY 785 ATGACGGTGAAGCTTGTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 844  
DB 9840 GTTATTTATATGATGATAATATATATATATATATATATATATATATATATAT 9899  
OY 845 AAGATATGTTAATGATGAGGGGTTGCGTATTT-TTATATTTAATCATATATAAA 903  
DB 9900 TATTTATTTGAATATATGATTTTATTTAAATTTATGATTTATTTAAATGATTTA 9959  
OY 904 TCAACATATATATGTTATTTGCTTTTATATAGTTCCTTTTAAAGAGATGATGACC 963  
DB 9960 TTATTTATTTTATTTTATTTTATTTTGAAGAAATATAGTTATTAAGATTAATTTGTTT 10019  
OY 964 TAAAGTGGCTAAATATGCGTAAATTTGCTATATATGCTATATGCTCAAGATGACT 1023  
DB 10020 TAAATTTATATAGATATGAGGTTAGTTTTCATTTTATATATATATATATATAGGTA 10079  
OY 1024 ATTGCGCAATTTGCAATATGCTCCTGCTATGCAACATATTTGATGATGCTGTTTGG 1083  
DB 10080 TTTTGTATATGATATTTGATGATATTTTAAATTTTATTTTATATTTTATTTTGA 10139  
OY 1084 CTGCAAAATATAAATTTTCTGTTGAATA 1115  
DB 10140 TTGCAATATATATATATGATATGATTAAGTA 10171

RESULT 2  
ABL32219  
ID ABL32219 standard; DNA; 11422 BP.

XX  
AC ABL32219;

XX  
DT 26-MAR-2002 (first entry)

XX  
DE Human immune system associated gene SEQ ID NO: 192.

XX  
KW Human: immune system disease; cytosine methylation; antiasthmatic;  
KW antidiabetic; antiasthmatic; antiasthmatic; antiasthmatic;  
KW antidiabetic; antiasthmatic; antiasthmatic; antiasthmatic;  
KW antidiabetic; antiasthmatic; antiasthmatic; antiasthmatic;  
KW antidiabetic; antiasthmatic; antiasthmatic; antiasthmatic;  
KW acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
KW gene; ds.

XX  
OS Homo sapiens.

XX  
PN WO200200928-A2.

XX  
PD 03-JAN-2002.

XX  
PF 02-JUL-2001; 2001WO-EP07537.

XX  
PR 30-JUN-2000; 2000DE-1032529.

XX  
PR 01-SEP-2000; 2000DE-1043826.

XX  
PA (EPIG-) EPIGENOMICS AG.

XX  
PI Olek A, Piepenbrock C, Berlin K;

XX  
DR WPI; 2002-130909/17.

XX  
PT Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -

XX  
PS Claim 1; SEQ ID NO 192; 32pp + Sequence Listing; German.

XX  
CC The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX  
SQ Sequence 11422 BP; 3913 A; 59 C; 1856 G; 5594 T; 0 other;

Query Match 2.0%; Score 72.8; DB 24; Length 11422;

Best Local Similarity 47.7%; Pred. No. 4.1e-07;

Matches 244; Conservative 0; Mismatches 267; Indels 1; Gaps 1;

OY 605 AATTATACATAAATATGATGCTATTATATGCTGATGCTATACACTTCATGCT 664  
DB 9660 AATTAAATTTTAAATGATGCTATGCTATGCTATATATAGATAATATAGTGTGT 9719  
OY 665 CTATACATGTAATATATGATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 724  
DB 9720 TAAATGCTTAT 9779  
OY 725 TTAATGCAAT 784  
DB 9780 TATTTTAT 9839  
OY 785 ATGACGGTGAAGCTTGTGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 844  
DB 9840 GTTATTTATATGATGATAATATATATATATATATATATATATATATATATAT 9899







CC immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours  
CC and cancer, particularly by determining status of cytosine methylation  
CC and/or by detecting single-nucleotide polymorphisms. Determination of  
CC individual methylation patterns may allow development of individualised  
CC therapies. The sequences given in records ABL92119-ABL9235 represent  
CC chemically pre-treated DNA fragments from genes associated with DNA  
CC repair, and their complements.  
CC Note: The sequence data for this patent is not represented in the  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.

SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 other;

Query Match	1.9%	Score 69.4	DB 24	Length 8093
Best Local Similarity	47.5%	Pred. No. 2.4e-06		
Matches 277; Conservative	0;	Mismatches 296;	Indels 10;	Gaps 2

Oy	527	ATGCTAGACGCGCATTTTCATTCATTTATTCACATATATTAACACGTCGTAAGAAGCGCTT	586
Db	6285	ATTGCTGTTTATTTTAAAAAATTTTGGTTAAAGTAATAATTTTGTTTAAAAATCATATTA	6344
Oy	587	TATTGGATGAATATTCGAATTTATTCACATTAATTAATGATGCTATTAATTAAGTGTATTT	646
Db	6345	GTATTAGATTGTTGGTAGCTTGAATATGATTTATTAAATTTGGAAATTCGCTATTA	6404
Oy	647	GGTATCAAGCTTCATGCTCTATACATGTAATATTAATTCGAGTTAGACCTTAATTCAGGT	706
Db	6405	TATTTTATATGTAAGATAGATATTAAGTAGAATATTTAAAGTATGTTTATTTATAGAT	6464
Oy	707	AAATTGCTATTTATATTAATTAATCTGAATATAATATGTAATGATTCCTTGTGTTATTTT	766
Db	6465	TGTAGTAAATTTTGATTTTATTTATTAAGATATTTGTTGTGTTAAATATGTAATTTTAA	6524
Oy	767	ATGTTTGTTTCATTTTAAATGAAGCGTAGCGTGTGCATTCATATTTTATGATGACAAAC	826
Db	6525	ATTTTGTATTATATGAAGAAGTAATTTTAAAGTTATATATGTAATAATTAATAATA	6584
Oy	827	ATCTTTGATGAGTATTTAAGATATTGTTAATGATGAGGCGTTGCGTATTTTTTAT	886
Db	6585	GCATTTTAAATTAATTTATGATTTTAAAGTATTAATTAATTTT - AAAAATTTTGTG	6642
Oy	887	ATTTAAATCAATAAATAATCAACAATATATGTAATTTTGCTGCTTTTATAGTCTCTTT	946
Db	6643	TTTAAGTTTGTTTTAAATTTATGAATAATTTTAAATTTTGCTGCTTTTATTAAGTAAATTT	6702
Oy	947	AAAGAGTAGAGTACCTTAAGGCTGCCCTAAATATGCGCTAAATTCGCTGCTTAATTT	1006
Db	6703	AAGGAATTAGCTTATGATTAAG-----AATTTAGGTGGAATTTGATTTGGGTAAATTT	6754
Oy	1007	CACCTCAAGATACACTATTGGCAAAATTCACAATATGTCACCTGCTATGAACAATATTT	1066
Db	6755	AAAATTAATGCTATTAAGAGTTTAAAAAATTAAGTGTGTTAGCGTTTATTAATAGAGT	6815
Oy	1067	AGTAGATGTTGTTTTCCTGCAAAAATTAATAATTTTTCGTTT 1109	
Db	6815	AGTAATTTGTTATTTTAGAGAGTGTGACGAATAATGCTTAGGGGT 6857	

RESULT 6  
ABK33974  
ID ABK33974 standard; DNA; 8093 BP.

AC ABK33974;

DT 18-JUN-2002 (first entry)

Human DNA for staging of Astrocytomas, complement, #29.

KW Human; ds; astrocytoma; cytostatic; staging; cysteine methylation; CpG;  
KW bisulphite; brain tissue; MALDI; ESI; electron spray mass spectrometry;;  
KW matrix assisted laser desorption/ionization mass spectrometry.

OS Homo sapiens.

XX WO200202808-A2.  
PN  
XX  
PD 10-JAN-2002.  
XX  
PE 02-JUL-2001; 2001WO-EP07538.  
XX  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPIG-) EPIGENOMICS AG.  
XX  
PI Olek A, Plegenbrock C, Berlin K;  
XX  
DR WPI: 2002-171649/22.  
XX  
XX  
PT Novel chemically modified genomic DNA sequences, useful in the  
PT characterisation, classification, differentiation, grading, staging,  
PT treatment and/or diagnosis of astrocytomas or predisposition to  
PT astrocytomas -  
XX  
XX Claim 1; SEQ ID No 58; 37pp; English.  
XX

The invention relates to a nucleic acid comprising a sequence (I) of at least 18 bases in length of a segment of chemically pre-treated genomic DNA which has any one of the sequences of (ABK33919-ABK34032) or its complement. Also included are an oligonucleotide or peptide nucleic acid (or set thereof) of at least 9 nucleotides which hybridises to (I), primers for (I), probes for detecting cytosine methylation or single-nucleotide polymorphisms (SNP) in (I), an array of oligomers or peptide nucleic acids for analysing diseases associated with the methylation states of the CpG dinucleotides of (I). The array is useful for determining genetic and/or epigenetic parameters, classification, differentiation, grading, staging, treatment and/or diagnosis of astrocytomas, or the predisposition to astrocytomas by analysing cytosine methylations, involves obtaining a biological sample containing genomic DNA, extracting the genomic DNA, converting cytosine bases which are unmethylated at the 5-position, in the genomic DNA sample, to uracil or another base which is dissimilar to cytosine in terms of hybridisation behaviour, by chemical treatment and amplifying chemically pre-treated genomic DNA fragments using the array and a polymerase, where the amplificates carry a detectable label. The method further involves identifying methylation status of one or more cytosine positions, and analysing methylation status of the cytosine positions by reference to one or more data sets. The genomic DNA is chemically treated by using a bisulphite, hydrogen sulphite or disulphite. The amplification step amplifies DNA which is of particular interest in astrocytoma or brain tissue, based on the specific genomic methylation status of brain tissues, as opposed to background DNA. The amplificates carry a fluorescent label or radionuclide, optionally, the labels of the amplificates are detachable molecule fragments having a typical mass which are detected in a mass spectrometer. The fragments of chemically pre-treated genomic DNA to be amplified, have a single positive or negative charge for a better detectability in the mass spectrometer. Preferably, the amplificates or fragments of the amplificates are detected by matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI). The present sequence is one of the chemically pre-treated reference DNA samples of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

SQ Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 other;

Query Match	1.9%	Score	69.4	DB	24	Length	8093
Best Local Similarity	47.5%	Pred. No.	2.4e-06				
Matches 277; Conservative	0	Mismatches	296			Indels	10; Gaps 2

QY 527 ATTCGTTACTGCATATTTCATTAAATTAACAGTCTCAAAAAGCGCTGT 586  
||| | | | | |||| | ||| |||| | |  
Db 6285 ATTGCTGTGTTTATTTTTAAAAAATTTTGTTTAAAGTAATAATTTTTGTTTAAAAATCATATNA 6344

```
QY 587 TATTGGATGATATCGAATTTATACATATAATGATGCTATTATTACTTGCCTGATT 646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6345 GTATTAGATTTTGTAGATGTTAGAAATGATTTATTAAATTTGGAATGTCGTATA 6404
QY 647 GGTATCAACTTTCCTGCTATACATGTAATATATTGCGAGTTAGACCTTAATTCAGGT 706
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6405 TATTATTATGCTAAGATAGATATTAAGTAAATATTTAAAGTAGTTTATTATAGAT 6464
QY 707 AATTGCTATTTTATTATTATATGTAATATATGTAATGCTGCTTTGCTGTTATTTT 766
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6465 TGTAGTAAATTTTGTATTATTATTATTAAGTAAATTTGTTGCTTAAATAGTAAATTTTAA 6524
QY 767 ATGTTGTTTCATTTTAAATGACCGTGAGCTTGTCATTCATTTATTTTATGATGACAAAC 826
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6525 ATTTTGTGTTATTTGAAAGGTAATTTTAAAGTTTATTATGTAATTTATTAATA 6584
QY 827 ATCTTGTATGAGATTTTAAAGATTTGTTAAATGATGAGGCTTGCTGTTATTTTAT 886
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6585 GCATTTTAAATTTATTTATATAGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 6642
QY 887 ATTTAAATCATTAATAATCAACATATATGTTATTTGCTGCTTTTATAGTCTTCTTT 946
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6643 TTAAAGTTTGTATTATTAATGAAATATTTTAATTTGCTGTTATTAATTTAGTAAATTT 6702
QY 947 AAAGAGGTAGATGACCTAAAGTGCCTAAATATGCGTAATTTGCCATTTGCTATATAT 1006
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6703 AAGGAATTTAGGTTATGATTTAG-----AATTTAGGTGAATTTGATTTGGGTATTT 6754
QY 1007 CACCTCAAGATACACTATTGCGCAATTTGACAAATATGTCCTTCGTATGAAACAATATT 1066
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6755 AAAATTAATGTTATTAAGATTTTAAAGTTTAAAGTTGCTGTTTAAATTTAGAGGT 6814
QY 1067 AGTAGATGTTGTTTGTGCTGCAAAATTAATTTTCTGCTT 1109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTTAGAGGCT 6857
```

```
RESULT 7
ABL49332
ID ABL49332 standard; DNA: 8093 BP.
XX
AC ABL49332;
XX
DT 01-MAY-2002 (first entry)
XX
DE Human polynucleotide associated with DNA replication SEQ ID NO 32.
XX
KW Human: cytosolic; neuroprotective; nootropic; immunostimulant;
KW gene therapy; gene regulation; DNA replication; CENPB; DNA2L; ATR; CHD1L;
KW ERCC3; SNRPB1; RAD50; Lig2; cytosine methylation; Ataxia telangiectasia;
KW ATR-X; Bloom's syndrome; tumour; cancer; methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN W0200177377-A2.
XX
PD 18-OCT-2001.
XX
PE 06-APR-2001; 2001WO-EP03971.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017471/02.
XX
PT New nucleic acid sequences from chemically modified genes associated
```

PT with DNA replication, useful for analysing cytosine methylations for  
PT diagnosis and therapy of diseases e.g. Ataxia telangiectasia  
PS Claim 1; SEQ ID NO 32; 23pp + Sequence Listing; English.

CC The invention relates to nucleic acid sequences comprising at least 18  
CC bases of a chemically pretreated gene associated with gene regulation,  
CC selected from 94 genes (ABL49301-ABL49394) and/or complementary sequences  
CC associated with DNA replication, CENPB, DNA2L, ATR, CHD1L, ERCC3, SNRPB,  
CC RAD50 and Lig2. The chemical pretreatment converts cytosine bases  
CC unmethylated at the 5-position to uracil or another base with  
CC hybridisation behaviour dissimilar to cytosine, to enable analysis of  
CC cytosine methylations. The DNA sequences and method are useful in the  
CC diagnosis of diseases (or predisposition to diseases) associated with DNA  
CC replication and in therapy of such diseases, by enabling analysis of the  
CC cytosine methylation patterns of such genes. They are especially useful  
CC in diagnosis and therapy of e.g. Ataxia telangiectasia, ATR-X, Bloom's  
CC syndrome, solid tumours and cancer.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification and was supposed to be available directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences. However, the sequence data did  
CC not correspond to that referred to in the specification. The present data  
CC is taken from EPO data for the patent.

XX Sequence 8093 BP; 2419 A; 175 C; 1689 G; 3810 T; 0 other;

XX Query Match 1.9%; Score 69.4; DB 24; Length 8093;

XX Best Local Similarity 47.5%; Pred. No. 2.4e-06;  
XX Matches 277; Conservative 0; Mismatches 296; Indels 10; Gaps 2;

```
QY 527 ATTCGTTACTGCTCATATTCATTAATTAATTCATTAATTAATTAATTAATTAATTAATTA 586
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6285 ATGTTGCTTTTATTTTAAATTTTGTTTTAAAGTAAATTTTGTTTTAAATTAATTAATTA 6344
QY 587 TATTGGATGATATCGAATTTATACATATAATGATGCTATTATTACTTGCCTGATT 646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6345 GTATTAGATTTTGTAGATGTTAGAAATGATTTATTTTAAATTTGGAATTTGCGTATA 6404
QY 647 GGTATCAACTTTCCTGCTATACATGTAATATATGTTGAGTGAACCTTAATTCAGAGT 706
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6405 TATTATTATGCTAAGATAGATATTAAGTAAGAAATTTAAAGATGTTTATTAATTTAGAT 6464
QY 707 AATTGCTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 766
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6465 TGTAGTAAATTTGTTATTTTAAAGATTAATTTGTTGCTTAAATTAATTTTAA 6524
QY 767 ATGTTGTTTCATTTTAAATGACCGTGAGCTTGTCATTCATTTTATGATGACAAAC 826
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6525 ATTTTGTGTTATTTGAAAGGTAATTTTAAAGTTTATTAATTAATTAATTTAATA 6584
QY 827 ATCTTGTATGAGATTTTAAAGATTTGTTAAATGATGAGGCTTGCTGTTATTTTAT 886
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6585 GCATTTTAAATTTATTTATAGATTTTAAAGTATTAATTTTAAATTTTAAATTTTAA 6642
QY 887 ATTTAAATCATTAATAATCAACATATATGTTATTTGCTGCTTTTATAGTCTTCTTT 946
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6643 TTAAAGTTTGTATTATTAATTTGAAATTAATTTTGTGTTTATTAATTTAGTAAATTT 6702
QY 947 AAAGAGGTAGATGACCTAAAGTGCCTAAATATGCGTAATTTGCCATTTGCTATATAT 1006
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6703 AAGGAATTTAGGTTATGATTTAG-----AATTTAGGTGAATTTGATTTGGGTATTT 6754
QY 1007 CACCTCAAGATACACTATTGCGCAATTTGACAAATATGTCCTTCGTATGAAACAATATT 1066
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6755 AAAATTAATGTTATTAAGATTTTAAAGTTTAAAGTTGCTGTTTAAATTTAGAGGT 6814
QY 1067 AGTAGATGTTGTTTGTGCTGCAAAATTAATTTTCTGCTT 1109
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6815 AGTAATTTGTTATTTTAGAGGATGTTGAGAAATGTTAGAGGCT 6857
```

RESULT 8  
ABL70350



5-position to uracil or another base with hybridisation behaviour  
dis-similar to cytosine, to enable analysis of cytosine methylations.  
The DNA sequences, oligomers (or sets/arrays) and method are  
useful in the diagnosis of diseases (or predisposition to diseases)  
associated with gene regulation and in therapy of such diseases, by  
enabling analysis of the cytosine methylation patterns of such genes  
kits are provided. They are especially useful in diagnosis  
and therapy of e.g. severe combined immunodeficiency disease, cardiac  
disorders, haemophilia, solid tumours and cancer, Werner syndrome,  
asthma, HDR syndrome, Seethre-Chotzen syndrome, renal disease,  
pre-eclampsia, graft versus-host disease. The present sequence is a  
sequence included in the sequence data for this specification and is  
associated with the human gene regulation-associated genes.  
Note: The sequence data for this patent did not form part  
of the printed specification, but was obtained in electronic  
format directly from WIPO at  
ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 5504 BP; 1511 A; 103 C; 1144 G; 2746 T; 0 other;

Query Match	1.98;	Score 69;	DB 24;	Length 5504;
PostgreSQL 9.1.10	1.98;	Score 69;	DB 24;	Length 5504;

Best Local Similarity 46.28; Pred. No. 2.6e-06;

Matches 265; Conservative 0; Mismatches 305; Indels 3; Gaps 1.

OY	615	TAATAAATGAAAGCATATATACCTGCGATATGCGTATCAACTTCATGCGTATACAGT	674
Db	1009	TTAATAATAGAGTGTGTTTGTGTAATTTTATAGGTTTTATTTATTCGTTTTTTTTTTTT	1068
OY	675	AATATATTTGAGATTTAGACCGTTAATTCAGGTAATTTGCTATTTAATATATATCTGAT	734
Db	1069	TTTATATTTAGCGT---TTTTTTTTTATATTTTTTTTTTATTTTTTTTATTTGTTT	1128
OY	735	AATATGTAATCGATTTGCTTTGTGGTATTTTATATGTTTCTTCATTTTTTAATGACGGTGA	794
Db	1126	TTTAGGATTAATTTTTTTTATGAGTAAATTTGATCGTATATATTTTTTTAGTCGTTTG	1185
OY	795	GCTTGCGATCTCATATTTTTATGANGACACACCTTGTATGAGATATTAAGATATATGT	854
Db	1186	TATTTTTTTTTTGAAGCTTTTTTATTTATAGAGCTTTTGTTTTTTTTATTTTAAAGTATTA	1245
OY	855	TAAATCATGAGGGGTTTGCCTGTATTTTTTATATTTAATCATATATTAATCAACAATATA	914
Db	1246	TGTGTTGCGGTGATTTTTTGATATATTAGTAAAGAAAAGATTTTAAAGATTTGTGAT	1305
OY	915	TGTTATTTTGTCTCTTTTATAGTGTCTTTTAAAGAGTAGAGATGACCTAAAGTCCGC	974
Db	1306	TTGATTAATTTTTTTTTTTTATTTATTTTTTTTTTTTTTTTTTTTGTATAGGTTGATATTTTT	1365
OY	975	TAAATATAGGCGTAAATTTGCCATGCTATTAATTCACCTCAACAATATCATATGGCAATT	1034
Db	1366	TAAATTTTATATTAATATAGCGCTTAACGCTTAATTAATTAATTAATTAAGTGTTAATTTATA	1425
OY	1035	GACAAATATGTCACCTCGATGTAAGAACAATATATGATAGTTGTTTTGCTGCACAAAATA	1094
Db	1426	GTTAATAGTTTTATATATTATTTAGTAGTTAAAGTATTTTATTTTAAATTTGTATAGTTAAAG	1485
OY	1095	AAAAATTTTCTGCTTGAATAAATCAAGCGCTCAGCGTTTCTTTATCTTAAATAACA	1154
Db	1486	AAGATAGTATATGTAATTTTATTTTATTTGCAAAATTAATGTTGATAGTAACTTTTATTAAG	1545
OY	1155	GGAATATGCGATGTAAGTATTAATGACACATAAG	1187
Db	1546	TTTAAAAAAAATGCTATTTATTTATAGCGTTAG	1578

## RESULT 10

ID ABQ67094 standard; DNA; 83391 BP

AC ABQ67094;

DT 28-AUG-2002 (first entry)

XX

DE Human angiogenesis associated polynucleotide SEQ ID NO 124

KW Human, angiogenesis; methylation; eye disease; glaucoma; tumour;  
KW Inflammation; rheumatoid arthritis; diabetic retinopathy; antiulcers;  
KW macular degeneration; inflammatory bowel disease; Crohn's disease;  
KW antidiabetic; antihypertensive; antidiabetic; antipsoriatic;  
KW antiarteriosclerotic; ds.

OS Homo sapiens

PN WO200246454-A2

PD 13-JUN-2002.

PF 06-DEC-2001; 2001WO-EP14320  
 YY

PR 06-DEC-2000; 2000DE-1061338

PA (EPiG-) EPIGENOMICS AG.

PI Schacht O;  
xx

DR WPI; 2002-500450/53.

PT New nucleic acid fragments from chemically treated  
PT angiogenesis-associated genes, useful for determining methylation  
PT status, e.g. in diagnosis or treatment of cancer -  
XX  
PS  
Claim 1; SEQ ID NO 124; 41pp + Sequence Listing; German.

CC The invention relates to a nucleic acid (I) comprising a segment of 18  
CC bases of chemically pretreated DNA of angiogenesis-associated genes (II)  
CC having sequences (ABQ66971-ABQ67178) or their complements. (1), also  
CC related oligomers, are used to evaluate the methylation status and/or  
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
CC neovascular glaucoma, solid tumours, inflammation, Rheumatoid arthritis,  
CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
CC Crohn's disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct/sequences](http://wipo.int/pub/published_pct/sequences).

SQ Sequence 83391 BP; 24547 A; 665 C; 16953 G; 41209 T; 17 other;

Query Match	1.98;	Score 68.2;	DB 24;	Length 83391;
-------------	-------	-------------	--------	---------------

Best Local Similarity 45.78; Pred. No. 1.1e-05;

Matches 363; Conservative 0; Mismatches 418; Indels 14; Gaps 3;

OY	515	TTTGGTCGTTAATCGATTACGCATCATTCATAATTAAATCAGTAATTAACAAGTCC	574
Db	56581	TATGATTAATTAATTAATTTTTTGATTAGTTTTTTTGGCTTTTATTTTAAATTAAGTTTTT	56640
OY	575	TAAAGGCGTGTATTGGATGAATAATCGAAATATACATATATAATGATGCAATTAAT	634
Db	56641	TTTATTTGTTTTTACTTTTTTATTTTATGTTAAATTAATATTGSGATTTTGTATTAAT	56700
OY	635	ACTTCGCTAATTTGATACATTCCTCATGCTCATACATGTAATATATATTCGAGTTAAC	694
Db	56701	AATATTAATTAATTTTAAATGTTATTTTAAGTTTTTATGTTGATTATTTGTTTT	56760
OY	695	TTAATTCAGGTAATTTGCTATTAATTAATTAATTAATTCGATAAATATGTAATGATTTG	754
Db	56761	TTGTTTTTTTAGTTAGTAATTTTACGTGGAAGTTATATAATTTTAAATATTTATAGG	56820
OY	755	GTCGTTAATTTTATGCTTTGTTTCATTTTAAATGACGGTGACCTTGTCATTCATATTTT	814
Db	56821	ATTTTACGTTAATGATTCGATTAATTTTATTTTATTTTATTTTATTAACGTTTTGTTTT	56880
OY	815	TATGATGACAAACCTTTGATGCAAGATTTTAAGAATTTGTTA-----ATGCATGAGG	866
Db	56881	TTTTGTTTTTTTATTAAGATTTTATTTATTTTAAATGTAATTTTATATAATGCTTTTT	56940

QY 867 GGTTCGCTGATTTTATTTATTAATCAATATAAATCAACATATATGTAATTTGCG 926  
 DB 56941 GATTTATTTTATTTTATTTTATTAATAATTTAATGTTAATAATTTTGTGTTTATG 57000  
 QY 927 TCTTTTATAGTCTTTTAAAGAGTAGATGACCTTAAAGTGCCCTAATATGCGCT 986  
 DB 57001 TTATATGATAGGAGCACTTATACGTTATGATTTTAAATTTTATTAATTTAT 57060  
 QY 987 AATTCGCATTCGTATTAATTCACCTCAAGATACACTATTTGGCAATTTGCAATATGTC 1046  
 DB 57061 AATTTTATAGGATGTTTAAATTTGTTGATTAATTTTAAATATTTTGTAGTTATTTT 57120  
 QY 1047 ACTTCCTA---TGAACACATATTTAGTAGATGTTGTTGCTGCAAAAATAAATTTT 1102  
 DB 57121 ATTTTATTTTATTTATATATGACGATTTATTTTATTTTATTTTATTTTAAATATTT 57180  
 QY 1103 TCTGCTGAATPACTCAAGCCCTAGCGTTTCTTATCTTAAATACAGAAATAG 1162  
 DB 57181 TTTTATTTTATTTTCGTTATGATTTTATTTTATTTTATTTTGGGAAATAGAGTATTA 57240  
 QY 1163 CCATTGCACTTAATGACACTTAAGCAATAGTACACCTTAACAGACAGACCTATGCC 1222  
 DB 57241 AAT--GAGTTGTTATTTATTTAATAAATTTATTTAGTTTATTTTATTTTATTTATG 57298  
 QY 1223 TTTGTCAGACATCAATGAGCACTTCTTAACCTGAGATGATTCGATCTGTG 1282  
 DB 57299 TTTATTTTATTTTATTTAGATTAGAGATTTATTTATTTTGAAGATTAATTTGATTTATA 57358  
 QY 1283 CGTTATCTTAACT 1297  
 DB 57359 TGTATTTAATAAT 57373

RESULT 11  
 AAS45516 standard; DNA; 6306 BP.

AAS45516;

18-DEC-2001 (first entry)

Chemically pretreated genomic DNA associated with cell cycle #111.

Cell cycle: human; CPG dinucleotide; cytosine methylation; HIV; aging;  
 human immunodeficiency virus; neurodegenerative disorder; solid tumour;  
 graft-versus-host disease; glomerular disease; Lewy body disease; cancer;  
 arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritis;  
 immunosuppressive; antitumour; cytosolic; antiarteriosclerotic; ds;  
 PCR primer.

Homo sapiens.

WO200168911-A2.

20-SEP-2001.

15-MAR-2001; 2001WO-EP02945.

15-MAR-2000; 2000DE-1013847.

06-APR-2000; 2000DE-1019058.

07-APR-2000; 2000DE-1019173.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIC-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI: 2001-602751/68.

Designing primers and probes for analysing diseases associated with  
 cytosine methylation state e.g. arthritis, cancer, aging,

PT arteriosclerosis comprising fragments of chemically modified genes  
 PT associated with cell cycle -  
 PS Claim 1: SEQ ID No 221; 28pp; English.

Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA  
 molecules associated with the cell cycle and specific PCR primers of the  
 invention. The sequences are useful for detecting the methylation state  
 of all CPG dinucleotides in a sequence and therefore for analysing  
 associated diseases. By analysing cytosine methylations in the pretreated  
 DNA, genetic and/or epigenetic parameters for the diagnosis and therapy  
 of existing diseases or the predisposition to specific diseases can be  
 ascertained. The parameters may be compared to another set of genetic  
 and/or epigenetic parameters, the differences serving as basis for  
 diagnosis and/or prognosis events which are disadvantageous to patients.  
 The sequences of the invention are useful for the diagnosis and therapy  
 of HIV infection, neurodegenerative disorders, graft-versus-host disease,  
 aging, glomerular disease, Lewy body disease, arthritis,  
 arteriosclerosis, solid tumours and cancers.

SO Sequence 6306 BP; 1822 A; 190 C; 1449 G; 2845 T; 0 other;

Query Match 1.9%; Score 68; DB 22; Length 6306;  
 Best Local Similarity 49.8%; Pred. No. 4,8e-06;  
 Matches 201; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

QY 586 TTATTTGATGAATTCGAAATTTATCACAATTAATGATGCTATTTACTGCTAT 645  
 DB 1745 TTTATTTTATTTATTTTCGTTATTTTATTCGTTATTTTATTTTATTTTATTTTAT 1804  
 QY 646 TGGTATCACTTCATGCTATACATGATTAATTTTTCGATTCAGCTTAATTCAGG 705  
 DB 1805 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1864  
 QY 706 TAATTTGCTATTTAATTTATTTATCTGAATTAATTAATGATTCGTTGCTGCTATTT 765  
 DB 1865 TTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1924  
 QY 766 TATGTTGTTTCATTTTATATGACGGTGCAGCTTGCAATTTTATATGACAA 825  
 DB 1925 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1984  
 QY 826 CATCTTGATGAAGTATTTAAGATTTGTTAATGATGAGGGTTCGCGATTTT-- 883  
 DB 1985 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 2044  
 QY 884 -TATATTAATCAATAAATCAACAAATATATGTTATTTGTCCTTTTATATGTTTC 942  
 DB 2045 ATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 2104  
 QY 943 TTTTAAAGAGTATGATGACCTAAGCTGCGCTAATATGCGCT 986  
 DB 2105 TTTTAAAGAGTATGATTTTTCATCGCAAAATTAAGACCT 2148

RESULT 12  
 ABK28458  
 ID ABK28458 standard; DNA; 6306 BP.

ABK28458;

23-APR-2002 (first entry)

DNA transcription associated complementary genomic DNA #166.

DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;  
 PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;  
 single nucleotide polymorphism; adenosine deaminase deficiency; cancer;  
 viral infection; Sezary syndrome; haematological disorder; tuberculosis;  
 immunological disorder; Werner syndrome; developmental disorder;  
 psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;  
 neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;  
 myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;



KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;  
KM polyglutamine disorder; solid tumour.  
XX  
OS Unidentified.  
XX  
PN WO200192565-A2.  
XX  
PD 06-DEC-2001.  
XX  
PF 06-APR-2001; 2001MO-EP03973.  
XX  
PR 06-APR-2000; 2000DE-1019058.  
PR 07-APR-2000; 2000DE-1019173.  
PR 30-JUN-2000; 2000DE-1032529.  
PR 01-SEP-2000; 2000DE-1043826.  
XX  
PA (EPiG-) EPIGENOMICS AG.  
XX  
PI Olek A, Piepenbrock C, Berlin K;  
XX  
DR WPI; 2002-090046/12.  
XX  
PT New nucleic acids or oligomers, useful for diagnosing or treating  
PT diseases associated with DNA transcription, e.g. immunological  
PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid  
PT tumours or cancer -  
XX  
PS Claim 1; SEQ ID NO 332; 32pp; English.  
XX  
XX The invention relates to a nucleic acid, which comprises a segment of the  
XX chemically pretreated DNA of genes associated with DNA transcription from  
XX one of 346 sequences, and an oligomer, in particular an oligonucleotide  
XX or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical  
XX to the chemically pretreated DNA of genes associated with DNA  
XX transcription. The set of oligomer probes are useful for detecting the  
XX cytosine methylation state and/or single nucleotide polymorphisms (SNPs)  
XX in a chemically pretreated genomic DNA. The nucleic acids are useful for  
XX diagnosing or treating diseases associated with DNA transcription  
XX (particularly with the methylation status), e.g. adenosine deaminase  
XX deficiency, viral infection, retroviral infection, Sezary syndrome,  
XX hematological disorders, immunological disorders, Werner syndrome,  
XX tubercolusis, developmental disorders, psoriasis, Rieger's syndrome,  
XX neurological disorders, neurodegenerative disorders, Waardenburg  
XX syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial  
XX infarction, hypertension, angiogenesis, erythropoiesis, congenital heart  
XX disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours  
XX or cancer. Sequences ABK28127-ABK28472 represent DNA transcription  
XX associated genomic DNA molecules of the invention.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification but was obtained in electronic format directly from the  
XX European Patent Office.  
SQ Sequence 6306 BP; 1822 A; 190 C; 1449 G; 2845 T; 0 other;

Query Match 1.9%; Score 68; DB 24; Length 6306;  
Best Local Similarity 49.8%; Pred. No. 4.8e-06;  
Matches 201; Conservative 0; Mismatches 200; Indels 3; Gaps 1;

QY 586 TTATTCGATGATATTCGAAATTAATTCACATATTAATTCATTCCTGCTAT 645  
DB 1745 TTATTTTATTTATTCGTTATTTATTCGTTATTTTATTTTATTTTATTTTAT 1804  
QY 646 TCGATATACCTTCATGCTCATATCATATATTTTCGATGATGACCTAATTCAGG 705  
DB 1805 TTTTATTTATTTTATTTTATTTTATTTTATTTTACGGCTTATTTTAAATTTTAT 1864  
QY 706 TAATTCCTCATATTAATTAATTCGAATATATATGATATGATGCTTTGGTTATTTT 765  
DB 1865 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1924  
QY 766 TAATTCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 825  
DB 1925 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1984

QY 826 CATCTTGTGATGAAGTATTAATTAAGATATTTGTAATGACATGAGGGCTTGCTGTAATTTT-- 883  
DB 1985 TTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2044  
QY 884 -TATATTAAATCAATTAATAAATCAACAAATTAATGATTTTGTCTTTTATAGCTTC 942  
DB 2045 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 2104  
QY 943 TTTTAAAGAGCTAGATGACCTAAGGCTCCCAATATATGCGT 986  
DB 2105 TTTTAAAGAGATAGTATTTTTCGATCGGAAATTAAGACGT 2148

RESULT 13  
ABL92249  
ID ABL92249 standard; DNA; 13131 BP.  
XX  
AC ABL92249;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Chemically treated DNA repair gene fragment complementary to#29.  
XX  
XX DNA repair; cytosine methylation; PMS2L1; PMS2L12; PMS2L2; PMS2L3;  
XX PMS2; L4; PMS2L5; PMS2L6; MGMT; MSH2; NUDP1; TDG; INPPL1; RFC4;  
XX DITL1; FANCB; XRCB; ataxia telangiectasia; agtng; Bloom's syndrome;  
XX Cockayne syndrome; Nijmegen breakage syndrome; Werner syndrome;  
XX immunodeficiency; trichothiodystrophy; Fanconi's anaemia; solid tumour;  
XX cancer; ds.  
XX  
XX Unidentified.  
XX  
XX WO200181622-A2.  
XX  
XX 01-NOV-2001.  
XX  
XX 06-APR-2001; 2001MO-EP03972.  
XX  
XX 06-APR-2000; 2000DE-1019058.  
XX 07-APR-2000; 2000DE-1019173.  
XX 30-JUN-2000; 2000DE-1032529.  
XX 01-SEP-2000; 2000DE-1043826.  
XX  
XX (EPiG-) EPIGENOMICS AG.  
XX  
XX Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-034446/04.  
XX  
XX New nucleic acid derived from genes associated with DNA repair, useful  
XX for diagnosis, e.g. of ataxia telangiectasia, by determination of  
XX cytosine methylation -  
XX  
XX  
XX Claim 1; SEQ ID NO 58; 25pp + sequence listing; English.  
XX  
XX The invention relates to nucleic acids containing a sequence of at least  
XX 18 nucleotides of chemically treated DNA of genes associated with DNA  
XX repair, and their complements. The invention also relates to nucleic  
XX acids comprising at least 18 base pairs of the chemically pretreated DNA  
XX of genes associated with DNA repair selected from PMS2L1, PMS2L12,  
XX PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, NUDP1, TDG, INPPL1,  
XX RFC4, DITL1, FANCB, or XRCB. Nucleic acids of the invention and related  
XX oligomers, are useful for diagnosis of diseases associated with gene  
XX repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,  
XX Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,  
XX immunodeficiency, trichothiodystrophy, Fanconi's anaemia, solid tumours  
XX and cancer, particularly by determining status of cytosine methylation  
XX and/or by detecting single-nucleotide polymorphisms. Determination of  
XX individual methylation patterns may allow development of individualised  
XX therapies. The sequences given in records ABL92192-ABL92353 represent  
XX chemically pre-treated DNA fragments from genes associated with DNA  
XX repair, and their complements.



CC Note: The sequence data for this patent is not represented in the  
CC specification, but is based on sequence information supplied by the  
CC European Patent Office.

XX Sequence 13131 BP; 4026 A; 221 C; 2455 G; 6425 T; 4 other;

Query Match 1.9%; Score 68; DB 24; Length 13131;  
Best Local Similarity 48.1%; Pred. No. 6.2e-06;  
Matches 223; Conservative 0; Mismatches 240; Indels 1; Gaps 1;

```
OY 543 ATTCATTAATTCAGTATTAATTAACAGCTTGAAGGCTGTTATTCGATGAATATTC 602
CC      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3635 ATTTTTCAGTACGATTTTATTTATTTATTTTCGTTTCGTTTATTAATTTATTTATTT 3694
OY 603 GAAATATCAGATTAATTTGATGCTATTTTACTGCTATTTGGATCAACTTCATC 662
DB 3695 TTTATATGTTAAATTAAGATTTTATTTTAAATTTTAAATTTGATTTTAAATATTT 3754
OY 663 CTCTATACATGTAATTAATTTTCGAGTTAGACCTTAATTCAGGTAATTTGCTATTTAAT 722
DB 3755 TTTTATTTTTCGTTTATTAAGATTAATGTAATTTATAGTAAATTTGTTATATTTT 3814
OY 723 TATTAATGATTAATTAATGATGATTCCTTTGCTTATTTTAACTTTGTTCAATTTT 782
DB 3815 ATTTATTTTATTTTATTTTATTTTGGTTTAAAGATTTTGTATTTTATTTATATTTT 3874
OY 783 TAATGAGGTCGATTCGATTCATATTTTATGATGACACATCTTTGATGAGATAT 842
DB 3875 TTTTATATGTTATGATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 3934
OY 843 TTAAGATATTTGTTAATGATGAGGCTTTCGCTATTTTATTAATTAATTAATTAATAA 902
DB 3935 TTTATTAATTTAATTAATTAATTAATTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 3993
OY 903 ATCAACATTAATGATTTTGTCTCTTTTAAAGCTTCTTTTAAAGAGCTGATGAC 962
DB 3994 GACTTTTAAAGCTTAAATTTTATTTTATTTATTAATTAATTAATTAATTAATTAATTA 4053
OY 963 CTAAAGCTGCTAAATTAATGCTTAATTCATTTGCTATTAATTT 1006
DB 4054 ATTAAGTCAGCTGATTTTAAAGCTTTGTTATTTTATTTATTTATTTATTTATTTATTT 4097
```

RESULT 14  
AAS46305  
AAS46305 standard; DNA; 8666 BP.

XX AAS46305;

XX 18-DEC-2001 (first entry)

XX Tumour suppressor gene derived chemically modified sequence #27.

XX Human; tumour suppressor gene; oncogene; antitumour; cytosolic;  
KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
OS cytosine methylation; ds.

XX Homo sapiens.

XX WO200168912-A2.

XX 20-SEP-2001.

XX 15-MAR-2001; 2001WO-EP02955.

XX 15-MAR-2000; 2000DE-1013847.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPig-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;  
XX WPI; 2001-602752/68.

XX Fragments of chemically modified genes associated with tumour suppressor  
PT genes and oncogenes, useful in designing primers and probes for  
PT analysing diseases associated with cytosine methylation state e.g.  
PT cancer

PS Claim 1; SEQ ID No 27; 27pp; English.

XX The invention relates to a nucleic acid comprising a sequence of 18  
CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with  
CC bisulphite, of genes associated with tumour suppression and  
CC oncogenes having a sequence taken from 536 (actually 533 since  
CC numbers 408, 458 and 500 are missing from the sequence listing) sequences  
CC (SS) and sequences complementary to (SS). The nucleic acid may be a  
CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may  
CC form part of a set of probes for detecting the cytosine methylation state  
CC and/or single nucleotide polymorphisms and also to be used in an  
CC array for analysing diseases associated with CpG dinucleotides e.g.  
CC cancers and tumours. The probes can also be used in a method for  
CC ascertaining genetic and/or epigenetic parameters for the diagnosis  
CC and/or therapy of existing diseases or the predisposition to specific  
CC diseases, by analysing cytosine methylations. The parameters may be  
CC compared to another set of genetic and/or epigenetic parameters, the  
CC differences serving as basis for diagnosis and/or prognosis events which  
CC are disadvantageous to patients. The present sequence is one of the  
CC 533 genomic sequences derived from tumour suppressor genes and  
CC oncogenes.

CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published-pct-sequences.

XX Sequence 8666 BP; 2444 A; 168 C; 1868 G; 4186 T; 0 other;

Query Match 1.9%; Score 66.6; DB 22; Length 8666;  
Best Local Similarity 44.6%; Pred. No. 1.2e-05;  
Matches 261; Conservative 0; Mismatches 324; Indels 0; Gaps 0;

```
OY 525 TAAATTCCTACTGCTCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 584
DB 30 TAAATTCCTACTGCTCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 89
OY 585 TTTATTCGATCATATTTGCAATTTATTCACATTAATTAATTAATTAATTAATTAATTT 644
DB 90 TTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTATTTT 149
OY 645 TTGCTATCACTTTGATGCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 704
DB 150 ATTAAGAGCTATTTTAATTTTACGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 209
OY 705 GTAATTTGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 764
DB 210 TTTTGTAAATTAATTAATTTTATTTATTAATTAATTAATTAATTAATTAATTAATTTT 269
OY 765 TTTATTTGCTTCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 824
DB 270 TTTATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 329
OY 825 ACATCTTTGATGAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 884
DB 330 GTATTTTGAAGTATTTTATTTTATTTTGAAGGATTTTAAATTTTAAATTTTAAATTTT 389
OY 885 ATATTAATCATATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTATTT 944
DB 390 ATTTTGTGTTGATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 449
OY 945 TTAAGAGCTAGATGACCTAAAGCTGCTTAATTAATTAATTAATTAATTTGCAATTTA 1004
DB 450 GGATATTTGTTTATTTTAAATTAATTAATTAATTAATTAATTTTATTTTATTTTATTTTATTT 509
```







```

: CITY: Research Triangle Park
: STATE: NO. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: CH 0016/97
: FILING DATE: 31-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 595:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 658 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: PAG1408RP
: US-08-998-416-595

```

```

Query Match      1.8%; Score 65; DB 4; Length 658;
Best Local Similarity 49.0%; Pred. No. 1.7e-07;
Matches 173; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

QY 594 TGAATATTCGAATTTATACATATTAATTCATGCTATTACTGCTGATTCGATTCATCA
    |||||
DB 210 TTTATATTTGTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT
    |||||
QY 654 ACTTTCGCTATACATTAATTAATTTGAGTTAGACCTTAATTCAGATTAATTTGT 713
    |||||
DB 270 TATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT
    |||||
QY 714 CTATTTAATTTATTCGATTAATTAATTCGATTCGCTTGGCTATTTTATTTATTTG 773
    |||||
DB 330 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 389
    |||||
QY 774 TTTCAATTTTAAAGCGGTGACCTTGTCATTCATTTTATTTTATTTATTTATTTATTTG 833
    |||||
DB 390 TTAACCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 449
    |||||
QY 834 ATGAAGATTTAAGATTTGTTTATTAATGATGAGGGCTTGGCTGATTTTATTTATTTAAT 893
    |||||
DB 450 ATTAATATCAATAGTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTAT 509
    |||||
QY 894 CATATTAATTAATCAATATTAATTTATTTGCTTTTATTAATGATTTCTTTT 946
    |||||
DB 510 GAACATATATATATGCTATTTCTATTTCTATTTATTTATTTATTTATTTATTTATTT 562
    |||||

```

```

RESULT 4
US-08-883-795A-36
: Sequence 36, Application US/08883795A
: Patent No. 5985607
: GENERAL INFORMATION:
: APPLICANT: Delcove, Genevieve
: TITLE OF INVENTION: Recombinant DNA Molecules and Expression
: NUMBER OF SEQUENCES: 39

```

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BERESKIN & PARR
: STREET: 40 King Street West
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5H 3Y2
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/883,795A
: FILING DATE: 27-JUN-1997
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Gravelle, Micheline
: REGISTRATION NUMBER: 40,261
: REFERENCE/DOCKET NUMBER: 7841-062
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 364-7311
: TELEFAX: (416) 361-1398
: INFORMATION FOR SEQ ID NO: 36:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 665 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: Rn 32
: US-08-883-795A-36

```

```

Query Match      1.6%; Score 59.2; DB 2; Length 665;
Best Local Similarity 45.8%; Pred. No. 5.4e-06;
Matches 242; Conservative 0; Mismatches 283; Indels 3; Gaps 1;

QY 515 TTTGGTCCTTAAATTCGTTACGCTCATATTTCAATTAATTCACATTTAAATTAACAGTTC 574
    |||||
DB 112 TTTTATATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 171
    |||||
QY 575 TAAAGCGCTGTTATTTGAGATATTCGAATTTATGACATTAATTAATTTGCTATTTAT 634
    |||||
DB 172 TTTTATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 231
    |||||
QY 635 ACTTGCCTGTTGCTATCACTTTCATGCTCTATACATGTAATTAATTTGAGATTACACC 694
    |||||
DB 232 TTTTATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 291
    |||||
QY 695 TTAATTCAGATTAATTTGCTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 754
    |||||
DB 292 TTTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 351
    |||||
QY 755 GTGGTATTTTGA--TGTGTTTCAATTTTAAATGACGCTTGTCATTCATTAAT 811
    |||||
DB 352 TTTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 411
    |||||
QY 812 TTTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 871
    |||||
DB 412 TTTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 471
    |||||
QY 872 GCGGTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 931
    |||||
DB 472 ATATTTACATATTTTATTAATTAATTAATTTATTTATTTATTTATTTATTTATTTAAT 531
    |||||
QY 932 TTTATAGTCTTTTAAAGAGTATGATGACCTTAAGGCTGCTTAAATTTATGCTTAAT 991
    |||||
DB 532 ATAAATTAATTAATTTATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTA 591
    |||||
QY 992 GCCATTTGCTTAATTTACACCTCAATTAATTAATTTGCAATTAATTTGCAAA 1039
    |||||

```









```

;
;   REGISTRATION NUMBER: 38,241
;   REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 919-541-8587
;   TELEFAX: 919-541-8689
;   INFORMATION FOR SEQ ID NO: 1137:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 636 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   ORIGINAL SOURCE:
;   ORGANISM: PAG1692RP
;   US-08-998-416-1137

Query Match      1.3%; Score 48.4; DB 4; Length 636;
Best Local Similarity 44.7%; Pred. No. 0.0034;
Matches 234; Conservative 0; Mismatches 286; Indels 4; Gaps 1;

QY 523 TTTAATTCGTCATGCTCATTCATTCATTTATTCACATTTAAATTAACAGCTTCAAAAGC 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 TTTAAACCTATTATTATTCATTTATTAATAATTAATTGATTGATTAAATCTTATTATA 134
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 583 TGTATTGATGATGAATATTCGAATATTCACATTAATTAATGATGATTTACTT 638
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 135 TATTTATTATATTTATTTACTATTCATTCATTTATTAATTTATTAATTAATAAATA 194
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 639 GCTGTATTGATCAACTTTCATGCTCTATACATGTAATTAATTTGAGTTGACCTTAA 698
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 195 TATTTAATATGAAATTAATTAATTTACTCTATGTTCAAAATTTTAAATAGTTATAAATATTA 254
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 699 TTCAAGGATATTTGCTATTATTATTATTAATCTCAATTAATTAATGATTCATGCTTGTG 758
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 255 TTGATATTAATTAATTTCTTAATAATAATTAATTAATTAATTAATTAATTAATTA 314
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 759 TTAATTTTATGTTGTTCTATTTTAAATGACGCTGACCTGCTCATTAATTTTATG 818
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 315 TTTATTAATTTGTTTATTAATAATATATTTTATTTATTAATAAGATTATTAATTTAAA 374
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 819 ATGACAAACATCTTGTATGAAGTATTTAAGATATTGTTAATGATGAGGGTTTGCGTGA 878
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 375 TATTGAATATTATTAATTTTATTAATATCAATTTTAAATAATTAATGATGATTATA 434
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 879 TTTTATATATTAATCATATATAATCAACAATATATAGTATTTTGCTTTTATAGT 938
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 TTTATTAATCTTTTATTAAGAAATTAATTAATAATTAATTTTAACCTTAATTTCTATTA 494
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 939 GTTCTTTTAAAGAGTAGAGTGAAGCTTAAGGCTGAATATATGCGGTAATATGCGCATTTG 998
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 495 TTAATTTTATTAATTTATTAATAATATATTAATTTATTTATTTATTTATTTATTTAAA 554
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 999 CTATTAATTCACCTCAAGATACACTATTGGCAAAATTGACAAATA 1042
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 555 TTAATTTATTAATTAATTAATTTTATCATTTATTAATTAATAATA 598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-09-426-290-1
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglund, Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345,2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
```

```

;
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (21181)...(21403)
;   NAME/KEY: CDS
;   LOCATION: (95252)...(95430)
;   NAME/KEY: CDS
;   LOCATION: (101753)...(101996)
;   NAME/KEY: CDS
;   LOCATION: (110324)...(110439)
;   NAME/KEY: CDS
;   LOCATION: (124058)...(124278)
;   NAME/KEY: CDS
;   LOCATION: (127009)...(127130)
;   NAME/KEY: CDS
;   LOCATION: (128910)...(129139)
;   US-09-426-290-1

Query Match      1.3%; Score 48; DB 4; Length 168575;
Best Local Similarity 46.7%; Pred. No. 0.049;
Matches 185; Conservative 0; Mismatches 210; Indels 1; Gaps 1;

QY 542 TATTCATTAATTCACATTAATTAATAACAGCTTCAAAAGCGCTGTTAATGGAATATT 601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108944 TATATGTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109003
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 602 CGAATTTATACATTAATTAATGATGCTATTTACTTGTGTATGCTATCAACTTTCA 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109004 ATATATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109063
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 661 TGTCTATACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 720
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109064 TATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA 109123
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 721 ATTAATTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 780
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109124 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109183
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 781 TTTAATGACGCTGAGCTTGCTATTTATTTTATGATGACAAACATCTTGATGAGT 840
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109184 TTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 109243
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 841 ATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 900
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109244 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 109303
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 901 AAATCAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 936
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 109304 ATGTAGGAAATCGAATTTATTTATGTAATTTATGTA 109339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-07-867-106-2/c
; Sequence 2, Application US/07867106
; Patent No. 5389526
; GENERAL INFORMATION:
; APPLICANT: Stade, Martin B
; APPLICANT: Chang, Andy C M
; APPLICANT: Williams, Keith L
; TITLE OF INVENTION: Improved Plasmid Vectors for Cellular
;   SLIME Moulds of the Genus Dictyostelium
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Woodcock Washburn Kurtz Maciewicz & NO. 5389526r19
;   STREET: One Liberty Place 46th Floor
;   CITY: Philadelphia
;   STATE: PA
;   COUNTRY: USA
;   ZIP: 19103
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

1 SOFTWARE: PatentIn Release #1.0, Version #1.25
2 CURRENT APPLICATION DATA:
3 APPLICATION NUMBER: US/07/867,106
4 FILING DATE: 19920625
5 PRIOR APPLICATION DATA:
6 APPLICATION NUMBER: AU PJ 7187
7 APPLICATION NUMBER: PCT/AU90/00530
8 FILING DATE: 02-NOV-1989
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Feeoney, Joanne Longo
11 REGISTRATION NUMBER: 35,134
12 REFERENCE/DOCKET NUMBER: RICE-0002
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: 215-568-3100
15 TELEFAX: 215-568-3439
16 INFORMATION FOR SEQ ID NO: 2:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 5852 base pairs
19 TYPE: NUCLEIC ACID
20 STRANDEDNESS: single
21 TOPOLOGY: linear
22 MOLECULE TYPE: DNA (genomic)
23 ANTI-SENSE: NO
24 FEATURE:
25 NAME/KEY: CDS
26 LOCATION: 2378..5038
27 FEATURE:
28 NAME/KEY: CDS
29 LOCATION: 2378..5038
30 US-07-867-106-2

```

```
US-08-998-416-937
: Sequence 937, Application US/08998416
: Patent No. 6239264
: GENERAL INFORMATION:
: APPLICANT: Philippsen, Peter
: APPLICANT: Pohlmann, Rainer
: APPLICANT: Steiner, Sabine
: APPLICANT: Mohr, Christine
: APPLICANT: Wendland, Jurgen
: APPLICANT: Knechtle, Philipp
: APPLICANT: Redischung, Corinne
: TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPII
: NUMBER OF SEQUENCES: 1152
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6239264artis Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: No. 6239264th Carolina
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/998,416
: FILING DATE: 24-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: CH 0016/97
: FILING DATE: 31-DEC-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Weigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8587
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 937:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 665 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: PAG1581RP
: US-08-998-416-937

Query Match      1.2%  Score 44.8;  DB 4;  Length 665;
Best Local Similarity 47.5%;  Pred. No. 0.03;
Matches 206;  Conservative 0;  Mismatches 217;  Indels 11;  Gaps 2;

QY  527 ATTCGTTACTGTCATATTCATTAATTCATTAATTAACAGTTCATAAAGCGTGT 586
DB  223 ATTTTAAATGTTAATAAATAAATAAGAAATTAAGAAATTAATTAATTAATAGTATT 282
QY  587 TATTGGATGAATTCGAATTAATTCACATTAATGAATGATGTAATTTACTGCTGTAAT 646
DB  283 TATTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 342
QY  647 GGTATCACTTCATGCTCTATACATGTAATATAT--TTGAGTTAGACCTTAATTCGAAG 704
DB  343 AATATTAAGTGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 402
QY  705 GTAATTTGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 755
DB  403 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 462
QY  756 TGGTATTTTATGTTGTTTCAATTTTAAATGACGGTGAAGCTTGACATTCATATTTTAT 815
```

```
DB  463 TAATAGTCTACCCCTTAATTTGATATTTACTACTAATAAATATTAATTAATTAAT 522
QY  816 ATGATGACAAACATCTTTGATGAAGTATTTAAGATATTTGATGCGGTTTGGCT 875
DB  523 ATTAAGAAATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 582
QY  876 GTATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 935
DB  583 TTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 642
QY  936 AGTGTCTTTTAA 949
DB  643 AATTTTGAATA 656

RESULT 15
US-09-419-459-1/c
: Sequence 1, Application US/09419459
: Patent No. 6222028
: GENERAL INFORMATION:
: APPLICANT: Liu, Jin-Hao
: APPLICANT: Cheng, Kuo-Joan
: APPLICANT: Tsai, Cheng-Fang
: APPLICANT: Chang, Chia-Chieh
: TITLE OF INVENTION: CELLULOSE ENZYMES
: FILE REFERENCE: 08919/037001
: CURRENT APPLICATION NUMBER: US/09/419,459
: CURRENT FILING DATE: 1999-10-15
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 2110
: TYPE: DNA
: ORGANISM: Pityomyces rhizinflata
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (3)...(1749)
: US-09-419-459-1

Query Match      1.2%  Score 44.8;  DB 4;  Length 2110;
Best Local Similarity 54.4%;  Pred. No. 0.05;
Matches 111;  Conservative 0;  Mismatches 92;  Indels 1;  Gaps 1;

QY  748 TTGCTTTGCTGTTATTTTATGTTTGTTCATTTTAAATGACGGTGACCTTGTCATTC 807
DB  2105 TTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2046
QY  808 TATTTTATGATGACACATCTTTGATGAAGTATTAAGATTTGTAATGATGAGGG 867
DB  2045 ATTTTATATATCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1986
QY  868 GTTGGCTGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 927
DB  1985 TTTTATTAATTAATCTTTTATTTGCTTTA-AGTAAGAAATTAATTAATTAATTTATTT 1927
QY  928 CTTTATTAATGCTGCTTTTAAAGA 951
DB  1926 CATTCATTAATTAATTTTATTA 1903

Search completed: March 2, 2003, 01:09:02
Job time : 1426 secs
```



GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 19:33:51 ; Search time 7520 Seconds  
(without alignments)  
11996.176 Million cell updates/sec

Title: US-09-915-706a-1

Perfect score: 3588  
Sequence: 1 gtcgactatcattcgtatgty.....gatnaattgagcgcttaa 3588

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_Main:\*

1: /cgn2\_6/ptodata/1/pna/US0600.COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US06.COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US07.COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US080.COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081.COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US082.COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US083.COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US084.COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US085.COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US086.COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US087.COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US088.COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US089.COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US090.COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US091.COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US092.COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US093.COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US094.COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US095A.COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US095B.COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US095C.COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US095D.COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US096A.COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US096B.COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US096C.COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US096D.COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US096E.COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US097A.COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US097B.COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US097C.COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US098A.COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US098B.COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US098C.COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US099A.COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US099B.COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US099C.COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US099D.COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US100A.COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US100B.COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US101A.COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US101B.COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US102A.COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US102B.COMB.seq.\*

Result				Query				SUMMARIES				Description			
No.	Score	Match	Length	DB	ID										
1	3587	100.0	3588	34	US-09-915-706A-1								Sequence 1, App11		
2	78.8	2.2	714	17	US-09-346-965A-909								Sequence 909, App1		
3	76.4	2.1	948	18	US-09-446-520-13069								Sequence 13069, A		
4	75	2.1	908	25	US-09-666-355A-6574								Sequence 6574, A		
5	73.8	2.1	529	23	US-09-619-643-13271								Sequence 13271, A		
6	73.8	2.1	529	25	US-09-654-617-341187								Sequence 341187, A		
7	73.8	2.1	529	27	US-09-849-016-341187								Sequence 341187, A		
8	73.6	2.1	529	58	US-09-145-485-9119								Sequence 9119, App		
9	73.2	2.1	810	26	US-09-666-355A-6400								Sequence 6400, App		
10	73.2	2.0	1776	38	US-10-015-127-2386								Sequence 2386, App		
11	72.8	2.0	432	33	US-09-865-419A-1334								Sequence 1334, App		
12	72.6	2.0	961	56	US-09-126-265-3039								Sequence 3039, App		
13	71.4	2.0	1077	38	US-10-015-127-2817								Sequence 2817, App		
14	71	2.0	589	32	US-09-446-520-11985								Sequence 11985, A		
15	71	2.0	589	64	US-09-202-214-11985								Sequence 11985, A		
16	70.8	2.0	504	29	US-09-739-449-4492								Sequence 4492, App		
17	70.8	2.0	504	31	US-09-803-110-4492								Sequence 4492, App		
18	70.8	2.0	1210	60	US-09-168-133-720								Sequence 720, App		
19	70.8	2.0	44417	19	US-09-514-000-363								Sequence 363, App		
20	70.8	2.0	254289	29	US-09-739-449-212								Sequence 212, App		
21	70.8	2.0	254289	31	US-09-803-110-212								Sequence 212, App		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

C 22	70.6	2.0	436	25	US-09-654-617-340109	Sequence 340109,
C 23	70.5	2.0	436	27	US-09-664-016-340109	Sequence 340109,
C 24	70.6	2.0	436	58	US-60-145-485-6606	Sequence 6606, App
C 25	70.4	2.0	510	17	US-09-304-517A-249382	Sequence 249382,
C 26	70.4	2.0	510	17	US-09-371-146A-249382	Sequence 249382,
C 27	70.4	2.0	510	18	US-09-440-687-8907	Sequence 8907, App
C 28	70.4	2.0	510	25	US-09-654-617-62906	Sequence 62906, App
C 29	70.4	2.0	510	27	US-09-664-016-62906	Sequence 62906, App
C 30	70.4	2.0	510	37	US-09-985-678-249382	Sequence 249382,
C 31	70.4	2.0	510	58	US-60-144-084-39263	Sequence 39263,
C 32	69.6	1.9	824	64	US-60-207-458-18907	Sequence 18907, App
C 33	69.4	1.9	824	64	US-09-873-402A-79376	Sequence 79376, App
C 34	69.4	1.9	599	64	US-60-209-830-51127	Sequence 51127, App
C 35	69.4	1.9	8093	41	US-10-172-086-56	Sequence 26, App
C 36	69.2	1.9	498	25	US-09-654-617-180433	Sequence 180433,
C 37	69.2	1.9	498	27	US-09-664-016-180433	Sequence 180433,
C 38	69.2	1.9	1498	31	US-09-806-708A-22	Sequence 22, App
C 39	69	1.9	700	19	US-09-527-433-909	Sequence 909, App
C 40	69	1.9	805	56	US-60-126-265-127	Sequence 127, App
C 41	69	1.9	5504	42	US-10-221-613-270	Sequence 270, App
C 42	68.8	1.9	576	26	US-09-666-355A-5638	Sequence 5638, App
C 43	68.8	1.9	799	18	US-09-404-520-13067	Sequence 13067, App
C 44	68.4	1.9	405	23	US-09-619-643-14065	Sequence 14065, App
C 45	68.4	1.9	612	34	US-09-902-540-1357	Sequence 1357, App

## ALIGNMENTS

```

RESULT 1
US-09-915-706A-1
: Sequence 1, Application US/09915706A
: GENERAL INFORMATION:
: APPLICANT: NELSON, DAVID R.
: TITLE OF INVENTION: A LIVE, AVIRULENT STRAIN OF V. ANGUILLARUM THAT
: TITLE OF INVENTION: PROTECTS FISH AGAINST INFECTION BY VIRULENT V.
: TITLE OF INVENTION: ANGUILLARUM
: FILE REFERENCE: 5112
: CURRENT APPLICATION NUMBER: US/09/915,706A
: CURRENT FILING DATE: 2001-07-26
: NUMBER OF SEQ ID NOS: 4
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 1
: LENGTH: 3588
: TYPE: DNA
: ORGANISM: Vibrio anguillarum
: FEATURE:
: NAME/KEY: modified_base
: LOCATION: (3572)
: OTHER INFORMATION: a, t, c, g, other or unknown
US-09-915-706A-1

```

Query Match	100.0%;	Score 3587;	DB 34;	Length 3588;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 3588;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	GTGACGTAATTGCAATGATGATGGCGTACAGCAAGTGGCCATCTTGTTGGTAAACAACG	60
Qy	1	GTGACGTAATTGCAATGATGATGGCGTACAGCAAGTGGCCATCTTGTTGGTAAACAACG <td>60</td>	60
Db	1	GTGACGTAATTGCAATGATGATGGCGTACAGCAAGTGGCCATCTTGTTGGTAAACAACG	60
Qy	61	TGTATATAAAGCTGGTGCGGTTTCATCAAGTTGAACACAAATACATGATTTTCCCACTT	120
Db	61	TGTATATAAAGCTGGTGCGGTTTCATCAAGTTGAACACAAATACATGATTTTCCCACTT	120
Qy	121	CCGGAAGGGAAGGTGAAATACCTTTTGAGATCAGCCCTGTTTACAGCTTTTCAATG	180
Db	121	CCGGAAGGGAAGGTGAAATACCTTTTGAGATCAGCCCTGTTTACAGCTTTTCAATG	180
Qy	181	ATCTTTTTCGTCGTACGTTTGGAAAAATCGACGACGTGGTTGTATTCACACAACGTA	240
Db	181	ATCTTTTTCGTCGTACGTTTGGAAAAATCGACGACGTGGTTGTATTCACACAACGTA	240
Qy	241	AGTGGATCCAATATCTCTAATTTGATAATAAAGCTGCTGTCCTGTCCTATATCCTGTG	300

Db	241	ACTGATCCATTATCTATTGTAATAAATGCTGCTTGTCTTTGCTATATCTGTG	300
Qy	301	AATTGCAGAGTGCATATATACCTGAAAAAAGCGTTTCCAGAAATCTAATTCGTAGAC	360
Db	301	AATTGCAGAGTGCATATATACCTGAAAAAAGCGTTTCCAGAAATCTAATTCGTAGAC	360
Qy	361	ACACAACAGCTTTACCTAGGTTTGGTATGCATGTCATGTTTGGCGGATGGAACG	420
Db	361	ACACAACAGCTTTACCTAGGTTTGGTATGCATGTCATGTTTGGCGGATGGAACG	420
Qy	421	GAAAATCAGACCCGCGGATAGCGTTTCTCTCGAATTAATGGGTGCAATTAACCTT	480
Db	421	GAAAATCAGACCCGCGGATAGCGTTTCTCTCGAATTAATGGGTGCAATTAACCTT	480
Qy	481	TTGCTATCTGAAGCTTAATGTTGGAGGAGCGGGTTGGTGCCTTAATTCGTTACTGCTC	540
Db	481	TTGCTATCTGAAGCTTAATGTTGGAGGAGCGGGTTGGTGCCTTAATTCGTTACTGCTC	540
Qy	541	ATATTCAAATTAATTCACATTTAAATAACGTTCTAAAGCGCTGTTTATTTGGATGATAT	600
Db	541	ATATTCAAATTAATTCACATTTAAATAACGTTCTAAAGCGCTGTTTATTTGGATGATAT	600
Qy	601	TGCAATTAATACAAATTAATTAATGATGCTATTAATCTGCTGATATGGATCACTTCA	660
Db	601	TGCAATTAATACAAATTAATTAATGATGCTATTAATCTGCTGATATGGATCACTTCA	660
Qy	661	TGCTCTATACATGTAATAATATTTTCGAGTTAGACCTTAATTCAGGTAATTTGTCTATTTA	720
Db	661	TGCTCTATACATGTAATAATATTTTCGAGTTAGACCTTAATTCAGGTAATTTGTCTATTTA	720
Qy	721	ATTATTAATCGAATAATATGTAATCGATGCTTGTGGTTATTTTATGTTTGTTCATT	780
Db	721	ATTATTAATCGAATAATATGTAATCGATGCTTGTGGTTATTTTATGTTTGTTCATT	780
Qy	781	TTTAATGAGCGGTGAGCTTGCGAATCAATTTTATGATGACCAACATCTTGGATGAAGT	840
Db	781	TTTAATGAGCGGTGAGCTTGCGAATCAATTTTATGATGACCAACATCTTGGATGAAGT	840
Qy	841	ATTTAAGATATTTGTTAATGCAATGATGAGGGGTTTGGCTATTTTATTAATCATATA	900
Db	841	ATTTAAGATATTTGTTAATGCAATGATGAGGGGTTTGGCTATTTTATTAATCATATA	900
Qy	901	AAATCAACAATATGTTATTTTGTGCTTTTATATAGTGTCTTTTAAAGAGTATGATG	960
Db	901	AAATCAACAATATGTTATTTTGTGCTTTTATATAGTGTCTTTTAAAGAGTATGATG	960
Qy	961	ACCTAAGGTGCTTAATTAATGAGGTAAATGTCATGCTATTAATTCACCTCAAGATATC	1020
Db	961	ACCTAAGGTGCTTAATTAATGAGGTAAATGTCATGCTATTAATTCACCTCAAGATATC	1020
Qy	1021	ACTATTTGCAAAATTCACAAATATGTCACCTCGATTAACCAATATTTAGATGTTGTTT	1080
Db	1021	ACTATTTGCAAAATTCACAAATATGTCACCTCGATTAACCAATATTTAGATGTTGTTT	1080
Qy	1081	TTGCTGCAAAATTAATAATTTTTCGTGTTGAATAATCACTCAAGGCTCTAGGCTTTCCCT	1140
Db	1081	TTGCTGCAAAATTAATAATTTTTCGTGTTGAATAATCACTCAAGGCTCTAGGCTTTCCCT	1140
Qy	1141	TATCTTAATAATACGGAATAATGAGATTGAAGTTAATGACCTTAAGCAAAATGATCAAC	1200
Db	1141	TATCTTAATAATACGGAATAATGAGATTGAAGTTAATGACCTTAAGCAAAATGATCAAC	1200
Qy	1201	TAAACAGACAGAGACCTATGCTCTTTGCAAGCATCAAAATGAGAACCTTTCTAAACCTC	1260
Db	1201	TAAACAGACAGAGACCTATGCTCTTTGCAAGCATCAAAATGAGAACCTTTCTAAACCTC	1260
Qy	1261	TGAGTATGATTCGATCTGTGGGCTTTATCTTAACCTGAAAAAAGTGGCTTTGCGCCAT	1320
Db	1261	TGAGTATGATTCGATCTGTGGGCTTTATCTTAACCTGAAAAAAGTGGCTTTGCGCCAT	1320
Qy	1321	TACGTAATGAATTAATGTGCGCAAACTGCGCTGCTAGCTAAGTCAAAAACCTTAGTG	1380
Db	1321	TACGTAATGAATTAATGTGCGCAAACTGCGCTGCTAGCTAAGTCAAAAACCTTAGTG	1380

Db 1321 TAGTAATGAATTTAATGTGCGCAAACTGCGTGTCTAAGCTAAGTCAAAACCCCTAGTG 1380  
Qy 1381 CTGACGAGAGATGCGTTCTCAAGAGGATGTCTAATTAATGGAAGATTCCTCTGACA 1440  
Db 1381 CTTGACGAGAGATGCGTTCTCAAGAGGATGTCTAATTAATGGAAGATTCCTCTGACA 1440  
Qy 1441 GTTTGTACGAACAGTTTTCAAAAACAACAGAGATATCGAGCTCATCTGTTGTG 1500  
Db 1441 GTTTGTACGAACAGTTTTCAAAAACAACAGAGATATCGAGCTCATCTGTTGTG 1500  
Qy 1501 CTGCTCAATTCCTCTCTGATACCAATTAAGAGTGTGGGAATAGCCTTGATGTTAG 1560  
Db 1501 CTGCTCAATTCCTCTCTGATACCAATTAAGAGTGTGGGAATAGCCTTGATGTTAG 1560  
Qy 1561 CGGATTAAGTGAAGAACACCTGGGATACCTCAACCTGTACTACAGTTGAAGCGCTCA 1620  
Db 1561 CGGATTAAGTGAAGAACACCTGGGATACCTCAACCTGTACTACAGTTGAAGCGCTCA 1620  
Qy 1621 AATCTGATGATGAATGAAGGCAAGAAAGAGAGCAAGCAGATGCGAAAGTTAAAGCATTTT 1680  
Db 1621 AATCTGATGATGAATGAAGGCAAGAAAGAGAGCAAGCAGATGCGAAAGTTAAAGCATTTT 1680  
Qy 1681 TCCAACTAGTCGGCGATAGCGAGAAAGCTCGATTCCTATGCGCGGTGCTGCACTGC 1740  
Db 1681 TCCAACTAGTCGGCGATAGCGAGAAAGCTCGATTCCTATGCGCGGTGCTGCACTGC 1740  
Qy 1741 CCTTAGTCGGGGAGATACGTTTTTTTACTTTTCAAAGTGCAGAGAGAAAAAGCGAAATCA 1800  
Db 1741 CCTTAGTCGGGGAGATACGTTTTTTTACTTTTCAAAGTGCAGAGAGAAAAAGCGAAATCA 1800  
Qy 1801 GCCAACTGAATCTATGCTTACGACACGCGTGGCGCAAGAGCGTTTGCAATTCATTTCA 1860  
Db 1801 GCCAACTGAATCTATGCTTACGACACGCGTGGCGCAAGAGCGTTTGCAATTCATTTCA 1860  
Qy 1861 AGATGAAAAAGCCAAACGTTGTGTCAACCAGTTAGATCGTTGTGACGCTTGTGAGCA 1920  
Db 1861 AGATGAAAAAGCCAAACGTTGTGTGTCAACCAGTTAGATCGTTGTGACGCTTGTGAGCA 1920  
Qy 1921 CTAGTGTCAATTCCTCTAGGAGATCAAGTACCAATCTCGGATTTGGCAAGTCACTGCTTA 1980  
Db 1921 CTAGTGTCAATTCCTCTAGGAGATCAAGTACCAATCTCGGATTTGGCAAGTCACTGCTTA 1980  
Qy 1981 CCCGTGTGAAAAAGCTTTGGTTCATCTAAGTGAATTAAGTACCAAGCGGAGG 2040  
Db 1981 CCCGTGTGAAAAAGCTTTGGTTCATCTAAGTGAATTAAGTACCAAGCGGAGG 2040  
Qy 2041 CCAAGACAGTAGAGCAAGAGGTTGCCAAAGTTCAAGTTCTGAAGGGAGGCTGCCAAGCC 2100  
Db 2041 CCAAGACAGTAGAGCAAGAGGTTGCCAAAGTTCAAGTTCTGAAGGGAGGCTGCCAAGCC 2100  
Qy 2101 AATAGATACAAAACATATAGAGGATATACGATGCGATCAGACAGGCTCAGACGTTAA 2160  
Db 2101 AATAGATACAAAACATATAGAGGATATACGATGCGATCAGACAGGCTCAGACGTTAA 2160  
Qy 2161 GCCAACTTACAGCAGAGAAACCTCTCTGAACCTGGGTAAATTAACAATATGAACGAG 2220  
Db 2161 GCCAACTTACAGCAGAGAAACCTCTCTGAACCTGGGTAAATTAACAATATGAACGAG 2220  
Qy 2221 ACTTAGCTTTCATTTTGTGAGAGAGTCTCTGATTAATTTTGGCGAGAGCGACGATA 2280  
Db 2221 ACTTAGCTTTCATTTTGTGAGAGAGTCTCTGATTAATTTTGGCGAGAGCGACGATA 2280  
Qy 2281 GCCCAATTTTCATTTTGTGAGAGAGGATGATGGGATATTTATCTTACTAGT 2340  
Db 2281 GCCCAATTTTCATTTTGTGAGAGAGGATGATGGGATATTTATCTTACTAGT 2340  
Qy 2341 TGTGTCGAGAAATGATGTCGGAACAAACGCTGACGCTCTTAGTACATTTTAAATCCG 2400  
Db 2341 TGTGTCGAGAAATGATGTCGGAACAAACGCTGACGCTCTTAGTACATTTTAAATCCG 2400  
Qy 2401 CCGGATTTGAATCATCTGATCAGGTTTGTGTCGCGGAGAGTGATACCAACGCTGGGCA 2460  
Db 2401 CCGGATTTGAATCATCTGATCAGGTTTGTGTCGCGGAGAGTGATACCAACGCTGGGCA 2460

Qy 2461 TTGAAGCCCCCAACACCTCAACGAGGCTTCCGTTTGGATCCGGAAGTTGAG 2520  
Db 2461 TTGAAGCCCCCAACACCTCAACGAGGCTTCCGTTTGGATCCGGAAGTTGAG 2520  
Qy 2521 AGCATGTATCTCAGACTTCCCTGTAGATACCCAATCAAGCAAGATCAAAAACCAAT 2580  
Db 2521 AGCATGTATCTCAGACTTCCCTGTAGATACCCAATCAAGCAAGATCAAAAACCAAT 2580  
Qy 2581 CATCCGCTAGCTGGGCTCTGAGTTGGTAATTTGTTTAAAAAATTAAGAAAAATCATGGC 2640  
Db 2581 CATCCGCTAGCTGGGCTCTGAGTTGGTAATTTGTTTAAAAAATTAAGAAAAATCATGGC 2640  
Qy 2641 AAGTATTTACATGGGTGAAGCGGCTTCAAGTTGAGGGCGACGACTATCGCTAGCT 2700  
Db 2641 AAGTATTTACATGGGTGAAGCGGCTTCAAGTTGAGGGCGACGACTATCGCTAGCT 2700  
Qy 2701 AGAAGCGCTGAAGTAAAAATGACGGTGTGTTGCAATCAACTTACTTTGGGGTGG 2760  
Db 2701 AGAAGCGCTGAAGTAAAAATGACGGTGTGTTGCAATCAACTTACTTTGGGGTGG 2760  
Qy 2761 CGCTCGTAACGTTGCTATGACATCGGTAAACGACCAATGCGGATTCAGGCATGTTGG 2820  
Db 2761 CGCTCGTAACGTTGCTATGACATCGGTAAACGACCAATGCGGATTCAGGCATGTTGG 2820  
Qy 2821 CGTAAGCGAAGTTAGCGTAAAGAGTCGATGCTGTGGAACACCTACTGCTCTTA 2880  
Db 2821 CGTAAGCGAAGTTAGCGTAAAGAGTCGATGCTGTGGAACACCTACTGCTCTTA 2880  
Qy 2881 TTTATTCACCCAGCTAAAGACGCTAAACCTGTTGAGTTGCAATTTACTAAGCCTTCTAA 2940  
Db 2881 TTTATTCACCCAGCTAAAGACGCTAAACCTGTTGAGTTGCAATTTACTAAGCCTTCTAA 2940  
Qy 2941 CGATGTCGAAGTGCAGACGTTTACTTCCAAAGTTAACTGAAAAACACAGTTAGTTTC 3000  
Db 2941 CGATGTCGAAGTGCAGACGTTTACTTCCAAAGTTAACTGAAAAACACAGTTAGTTTC 3000  
Qy 3001 TTTCAAGGTGAGCGGAGCTGACGAGTCAACCGTACGAGAGGCTATCTCTTCTTACAC 3060  
Db 3001 TTTCAAGGTGAGCGGAGCTGACGAGTCAACCGTACGAGAGGCTATCTCTTCTTACAC 3060  
Qy 3061 TTTCTATTTCTCAGAGCATCATATGAGAAAGAGTGTGTAATACAAAGCGGTGTGT 3120  
Db 3061 TTTCTATTTCTCAGAGCATCATATGAGAAAGAGTGTGTAATACAAAGCGGTGTGT 3120  
Qy 3121 TGTGACTTACGACCTACCGACCGGAAAAATGACTTGTGTAAGTAATTTCTTCAATTAGAC 3180  
Db 3121 TGTGACTTACGACCTACCGACCGGAAAAATGACTTGTGTAAGTAATTTCTTCAATTAGAC 3180  
Qy 3181 ATGGCAGTTAATTTGGCATGCTATTTCAATGAATATCTCAATTTTGAAGACACCGTTATGGC 3240  
Db 3181 ATGGCAGTTAATTTGGCATGCTATTTCAATGAATATCTCAATTTTGAAGACACCGTTATGGC 3240  
Qy 3241 ATTGAACCTCAACATTAAGCGCTTGTAGTAAGAACCGGTGACATCACTATGACGTTGA 3300  
Db 3241 ATTGAACCTCAACATTAAGCGCTTGTAGTAAGAACCGGTGACATCACTATGACGTTGA 3300  
Qy 3301 AACGAATGCGCGCTTAAAGAGAGAGCTGCTTTGTTGTGGCTCATTTGCGACTT 3360  
Db 3301 AACGAATGCGCGCTTAAAGAGAGAGCTGCTTTGTTGTGGCTCATTTGCGACTT 3360  
Qy 3361 TTTGAGGACACAAACCAATTCAGAAAAAGTTGATTTGAAGAGGAGGATTCACGGGTAT 3420  
Db 3361 TTTGAGGACACAAACCAATTCAGAAAAAGTTGATTTGAAGAGGAGGATTCACGGGTAT 3420  
Qy 3421 CGATTAAGACAACTTTCGATACAGTATGAGGGCAAAATTCACCGCGTCTTTCGTAAGGT 3480  
Db 3421 CGATTAAGACAACTTTCGATACAGTATGAGGGCAAAATTCACCGCGTCTTTCGTAAGGT 3480  
Qy 3481 TGATTAACAAGCTTTCATATGATGATACCACTTTGAAGTGAACCTTCCGCTTGGAT 3540  
Db 3481 TGATTAACAAGCTTTCATATGATGATACCACTTTGAAGTGAACCTTCCGCTTGGAT 3540













```
; CURRENT FILING DATE: 2001-05-07
; PRIOR APPLICATION NUMBER: US 60/202,214
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/816,660
; PRIOR FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 30131
; SEQ ID NO 11985
; LENGTH: 589
; TYPE: DNA
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(589)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: uc-zmroteosinle051f12a1
US-09-849-526A-11985
```

```
Query Match      2.0%; Score 71; DB 32; Length 589;
Best Local Similarity 46.5%; Pred. No. 9.1e-05;
Matches 227; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
```

```
QY 457 ATTAATTCGCGTGAACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || || ||
DB 517 ATTAATTCGCGTGAACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || || ||
QY 517 TGGTCTTAATTCGTTACTGCTCATATTCGAATTAATTCACATTAATTAACAGTTCTA 576
  || || || || || || || || || || || || || || || || || || || || || ||
DB 457 TATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAA 398
  || || || || || || || || || || || || || || || || || || || || || ||
QY 577 AAGGCGCTTTATTCGATGATGATATTCGAATTAATTCACATTAATTCATGCTATTTAC 636
  || || || || || || || || || || || || || || || || || || || || || ||
DB 397 TATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAA 338
  || || || || || || || || || || || || || || || || || || || || || ||
QY 637 TTGCTGATTCGATCAACTTCATGCTCATATTCGAATTAATTCATGCTATTTACGAGTTAG 696
  || || || || || || || || || || || || || || || || || || || || || ||
DB 337 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 278
  || || || || || || || || || || || || || || || || || || || || || ||
QY 697 AATTCAGGTAATTTGCTATTTAATTAATTCGAATTAATTCGAATTAATTCGAATTCGTTGT 756
  || || || || || || || || || || || || || || || || || || || || || ||
DB 277 TATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 218
  || || || || || || || || || || || || || || || || || || || || || ||
QY 757 GGTATTTTATTTGTTGTTTCATTTTAATGACGGTGAGCTTGTGCTATCATATTTTAA 816
  || || || || || || || || || || || || || || || || || || || || || ||
DB 217 ATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA 158
  || || || || || || || || || || || || || || || || || || || || || ||
QY 817 TGATGACACATCTTCGATGAGTATTTAAGATATTTGTTAATGACGAGGCTTGGCTG 876
  || || || || || || || || || || || || || || || || || || || || || ||
DB 157 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 98
  || || || || || || || || || || || || || || || || || || || || || ||
QY 877 TATTTTATTTAATTAATCAATTAATAATCAACATATATGCTATTTGTCCTTTTATA 936
  || || || || || || || || || || || || || || || || || || || || || ||
DB 97 TATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 38
  || || || || || || || || || || || || || || || || || || || || || ||
QY 937 GTGTCTT 944
  || || || || || || || || || || || || || || || || || || || || || ||
DB 37 GTTTATTT 30
  || || || || || || || || || || || || || || || || || || || || || ||
```

```
RESULT 15
US-60-202-214-11985/c
; Sequence 11985, Application US/60202214
; GENERAL INFORMATION:
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Shukla, Hridayabhiraaj
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51930)A
; CURRENT APPLICATION NUMBER: US/60/202,214
; CURRENT FILING DATE: 2000-05-08
```

```
; NUMBER OF SEQ ID NOS: 29880
; SEQ ID NO 11985
; LENGTH: 589
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: uc-zmroteosinle051f12a1
US-60-202-214-11985
```

```
Query Match      2.0%; Score 71; DB 64; Length 589;
Best Local Similarity 46.5%; Pred. No. 9.1e-05;
Matches 227; Conservative 0; Mismatches 261; Indels 0; Gaps 0;
```

```
QY 457 ATTAATTCGCGTGAACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || || ||
DB 517 ATTAATTCGCGTGAACAATATACCTTTGCTATCTGAAAGCTTAATGCGAGGAGCGGCTT 516
  || || || || || || || || || || || || || || || || || || || || || ||
QY 517 TGGTCTTAATTCGTTACTGCTCATATTCGAATTAATTCACATTAATTAACAGTTCTA 576
  || || || || || || || || || || || || || || || || || || || || || ||
DB 457 TATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAA 398
  || || || || || || || || || || || || || || || || || || || || || ||
QY 577 AAGGCGCTTTATTCGATGATGATATTCGAATTAATTCACATTAATTCATGCTATTTAC 636
  || || || || || || || || || || || || || || || || || || || || || ||
DB 397 TATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAA 338
  || || || || || || || || || || || || || || || || || || || || || ||
QY 637 TTGCTGATTCGATCAACTTCATGCTCATATTCGAATTAATTCATGCTATTTACGAGTTAG 696
  || || || || || || || || || || || || || || || || || || || || || ||
DB 337 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 278
  || || || || || || || || || || || || || || || || || || || || || ||
QY 697 AATTCAGGTAATTTGCTATTTAATTAATTCGAATTAATTCGAATTAATTCGAATTCGTTGT 756
  || || || || || || || || || || || || || || || || || || || || || ||
DB 277 TATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 218
  || || || || || || || || || || || || || || || || || || || || || ||
QY 757 GGTATTTTATTTGTTGTTTCATTTTAATGACGGTGAGCTTGTGCTATCATATTTTAA 816
  || || || || || || || || || || || || || || || || || || || || || ||
DB 217 ATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA 158
  || || || || || || || || || || || || || || || || || || || || || ||
QY 817 TGATGACACATCTTCGATGAGTATTTAAGATATTTGTTAATGACGAGGCTTGGCTG 876
  || || || || || || || || || || || || || || || || || || || || || ||
DB 157 TTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 98
  || || || || || || || || || || || || || || || || || || || || || ||
QY 877 TATTTTATTTAATTAATCAATTAATAATCAACATATATGCTATTTGTCCTTTTATA 936
  || || || || || || || || || || || || || || || || || || || || || ||
DB 97 TATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT 38
  || || || || || || || || || || || || || || || || || || || || || ||
QY 937 GTGTCTT 944
  || || || || || || || || || || || || || || || || || || || || || ||
DB 37 GTTTATTT 30
  || || || || || || || || || || || || || || || || || || || || || ||
```

```
Search completed: March 2, 2003, 03:23:35
Job time : 7550 secs
```



GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 1, 2003, 11:54:29 ; Search time 4780 Seconds  
(without alignments)  
12156.780 Million cell updates/sec

Title: US-09-915-706a-1

Perfect score: 3588

Sequence: 1 gtgcactatcgtcattgctgcttgccttaacgacgctttaa 3588

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: em\_estba:\*  
10: em\_estba:\*  
11: em\_estba:\*  
12: em\_estba:\*  
13: em\_estba:\*  
14: em\_estba:\*  
15: em\_estba:\*  
16: em\_estba:\*  
17: em\_estba:\*  
18: em\_estba:\*  
19: em\_estba:\*  
20: em\_estba:\*  
21: em\_estba:\*  
22: em\_estba:\*  
23: em\_estba:\*  
24: em\_estba:\*  
25: em\_estba:\*  
26: em\_estba:\*  
27: em\_estba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	91	2.5	1101	17	CNS00EVL
2	82	2.3	1094	17	CNS012FZ
3	78.8	2.2	714	17	AQ050116
4	76.6	2.1	1101	17	CNS0039G
5	76.6	2.1	1101	17	CNS00H04
6	75.2	2.1	781	17	CNS009DO

Result No.	Score	Query Match	Length	ID	Description
7	74.4	2.1	1101	17	CNS0039G
8	74.2	2.1	576	17	CNS035N7
9	73.8	2.1	1101	17	CNS012FZ
10	73.4	2.0	996	17	CNS00P0H
11	73.2	2.0	870	17	CNS009XJ
12	72.8	2.0	991	17	CNS0012D
13	72.8	2.0	1084	17	CNS006EL
14	72.8	2.0	1200	17	CNS016EL
15	72.6	2.0	614	17	CNS0152H
16	72.4	2.0	1539	17	A0897537
17	72.2	2.0	897	17	CNS07ABZ
18	72.2	2.0	945	13	BM358135
19	72.2	2.0	945	17	CNS04D0K
20	71.6	2.0	900	17	CNS015J4
21	71.4	2.0	1092	17	CNS020K7
22	71.2	2.0	662	17	CNS0419K
23	71.2	2.0	759	17	CNS060XV
24	71.2	2.0	927	17	A2547739
25	71.2	2.0	1001	17	CNS0155H
26	71.2	2.0	1036	17	CNS00599
27	70.6	2.0	609	17	CNS025K2
28	70.6	2.0	1009	17	CNS00587
29	70.4	2.0	876	17	CNS009G1
30	70.4	2.0	1101	17	CNS017V2
31	70.2	2.0	1101	17	CNS001FB
32	70.2	2.0	1101	17	CNS00EVL
33	70.2	2.0	1169	17	CNS006KH
34	70.2	2.0	900	17	A2678376
35	69.8	1.9	783	17	CNS00A1S
36	69.6	1.9	767	17	CNS00A0X
37	69.4	1.9	717	17	AG154664
38	69.4	1.9	842	17	CNS0100N
39	69.4	1.9	1135	17	CNS033G0
40	69.2	1.9	1190	17	CNS020N7
41	69	1.9	1101	17	CNS01348
42	69	1.9	1101	17	CNS0153V
43	68.8	1.9	1200	17	CNS016C0
44	68.8	1.9	878	17	CNS0028X
45	68.8	1.9	1101	17	CNS0021J

## ALIGNMENTS

RESULT 1  
CNS00EVL/c 1101 bp DNA linear GSS 04-JUN-1999  
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:  
BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.  
ACCESSION AL069706 GI:4949849  
VERSION AL069706.1  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
REFERENCE  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial







```
Db 1091 WDRFTKDDMDWTWTKWTKDRADRRWAGDADRWAMDGATWWTATWMMWMMWMTWD 1032
Qy 674 TAATATATTCAGTAGACCTTAATCAAGTAATTTGCTATTAATTAATTAATTCGAA 733
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 1031 TWWKMMWMAAATKTDIAWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTWMTW 972
Qy 734 TAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 793
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 971 KDKKRDGDDGDDGKKKKKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 912
Qy 794 AGCTTGATCATCATATTTTATGATGACACATCTTGATGATGATGATGATGATGATG 853
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 911 DDDGKGADADDDTGTGATDGDGDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDKDK 852
Qy 854 TTAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 911
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 851 DAADDDWADDDWADDDWADDDWADDDWADDDWADDDWADDDWADDDWADDDWADDD 792
Qy 912 ATATGATATTTTGTGCTTTTATAGTGTCTTTTAAAGAGATGATGATGATGATGATG 971
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 791 AADRDADATTTTWTTTTTRDTDDMKMTDTWTWMAADRTWMDRDDDDDRDRAAGTAK 732
Qy 972 GCCCAATATGCGCAATATGCGCAATATGCGCAATATGCGCAATATGCGCAATATG 1031
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 731 RTWRMRMRRTWRMRMRRTWRMRMRRTWRMRMRRTWRMRMRRTWRMRMRRTWRMR 672
Qy 1032 ATATGACAAATATGACAAATATGACAAATATGACAAATATGACAAATATGACAA 1091
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 671 WMAADWMTTDTDTDDDKDRRKRARRRRRTTARAAMWMTWMTWMTWMTWMTWMTW 612
Qy 1092 ATAAATATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1151
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 611 RMDRMAADTWTDAKADBDWMAKARAWRARARARARARARARARARARARARAR 552
Qy 1152 AGAGAAATACAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 1211
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 551 AWWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 492
Qy 1212 GAACCTATGCTTGTCAAGATGATGATGATGATGATGATGATGATGATGATGAT 1271
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 491 AAAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 432
Qy 1272 TCGATCTGTGCGTTTATCTTAACTGAAAGGAGGAGGAGGAGGAGGAGGAGG 1318
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 431 WAATTAATTTTWTWMAAAAAAAMWMTATATATKCCCCCCYCCC 385

RESULT 5
CNS00H04 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC:
DEFINITION BACR35018 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL074046.1 GI:4953725
VERSION AL074046.1 GI:4953725
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT
- Web : www.genoscope.cns.fr
- Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
```

```
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
pl and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1. 1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR35018"
/clone_11b="RPCI-98"
/note="end : 17"
location/Qualifiers

BASE COUNT 168 a 22 c 92 g 398 t 421 others
ORIGIN

Query Match 2.1%; Score 76.6; DB 17; Length 1101;
Best Local Similarity 34.7%; Pred. No. 5.6e-07;
Matches 191; Conservative 11; Mismatches 348; Indels 0; Gaps 0;

Qy 628 TATTATTACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 687
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 48 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 107
Qy 688 TTACACCTAATCAAGGATATTTGCTATTTATTTATTTATTTATTTATTTATTTAT 747
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 108 TTTGTDGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 167
Qy 748 TTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 807
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 168 TTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNT 227
Qy 808 TATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 867
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 228 TTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTT 287
Qy 868 GTTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 927
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 288 NTNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNNNTNNNN 347
Qy 928 CTTTATAGTGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 987
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 348 TTTTATTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTTNTTT 407
Qy 988 AATGCAATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1047
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 408 NNATTTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 467
Qy 1048 CTTGCTATGAAACATTTTACTAGATGATGATGATGATGATGATGATGATGATG 1107
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 468 NTATANNNTNNNTNTTACANNANNAANNNNNNNNNNNNNNNNNNNNNNNNNNT 527
Qy 1108 TTGAATTAACCAAGGCTAGGCTTGTCTTATCTTAAATACAGAAATAGCAT 1167
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 528 TTTAAANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNT 587
Qy 1168 GAAGTAAAT 1177
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
Db 588 TTTNTTTNTT 597

RESULT 6
CNS009DO/c 781 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19009 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053444
VERSION AL053444.1 GI:4934889
KEYWORDS GSS.
```

SOURCE ORGANISM	REFERENCE
Drosophila melanogaster.	Genoscope.
Drosophila melanogaster	Direct Submission
Eukariota: Metazoa: Arthropoda: Hexapoda: Insecta: Pterygota:	Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage
Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:	
Ephydroidea: Drosophilidae: Drosophila.	
1 (bases 1 to 781)	
Genoscope.	
Direct Submission	
Submitted (02-JUN-1999)	

**COMMENT**

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see [http://www.fruitfly.org/The\\_BDGP\\_Drosophila\\_melanogaster\\_BAC\\_library](http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library) was prepared by Kazutoyo Oosogawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

```

source
1..781
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19003"
/clone_1lb="RPC1-98"
/note="end : TE13"
440 a 71 c 6 g 189 t 75 others
BASE COUNT
ORIGIN

```

Query Match	2.1%	Score 75.2	DB 17	Length 781
Best Local Similarity	46.0%	Pred. No. 1.1e-06		
Matches 203; Conservative	21;	Mismatches 215;	Indels 2;	Gaps 1.

574 TTTGGTCTTAATTCGTTACTGCTCATATTCAATTAATTCACCTATTAAATAACAGTTC

**Dd**      439 TTTATTCTTAAATTAATAAATTTAAATTTTAAATTTTAA  
              380

575 TAAAGGCTGTTATTGCGATGATATTGCAATTATCACAATAATTGATGCTATTATT 634

[illegible]

635 ACTTGCTGTAATGGTATCAACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACC 694

Db 319 TTTTWTATTTTTTAATTTTTTAAATTAATTAATTTTAAAT 260

695 TTAATTCAAGGTAATTGCTCTATTAAATTATCTGAATAATATGTAATCGATTGCTTT 754

Db 259 TAATTTTAA--TAATWTTTWTATTATTATATTTTAAATTTTTTTAAW 202

QY 755 GTGGTATTTTATGTTTCTTCATTTTTAAATGACGGTGAGCTTGTGCATTCAATATTTT 814

Db 201 TAATTTWTTTTTTTTTTTTTTTTTTTTTTTTTTTTTATAATTTC 142.

815 TATGATGACAACATCTTTGATGAGTATTTAAGATATTGTTAATGCATGAGGGTTGCG 874

Db 141 TTTTATATTTTATTAATTTTWTATTTNATATAATTAATTTTAA 82

875 TGTATTTTATATTAATCATTAATAAATCAACAATATATGTTATTTTGTGCTTTTAA 934

Db           81 TTTTWWTTTTTAATTTAAAWTTTTTATTATAITAAWTTTTTY 22

QY 935 TAGTGTCTTTTAAGAGGTA 955

Db 21 TAATTTATTATTATTTA 1

Accession	Definition	LOCUS	CNS00396	1101 bp	DNA	linear	GSS 03-JUN-1999
CNS00396	Drosophila melanogaster genome survey sequence TE13 end of BAC # BACR08K10 of RP11-38 library from Drosophila melanogaster (fruit fly) genomic survey sequence.						
U007023							

VERSION	KEYWORDS
AL063921.1	GI:4941778
SS.	
SOURCE	ORGANISM
Drosophila melanogaster.	
Drosophila melanogaster	
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	
1 (bases 1 to 1101)	

**JOURNAL** Submitted (02-JUN-1999) Genoscope - Centre National de Séquençage : BP 191 91006 Evry cedex - FRANCE (E-mail : [segre@genoscope.cns.fr](mailto:segre@genoscope.cns.fr) - Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr))

**COMMENT** Determination of this BAC-end sequence was carried out as part of a

The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoso in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RP11-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES	source	Location/Qualifiers
	1. .1101	
	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/clone="BACR08K10"	
	/clone_1id="RPC1-98"	
	/note="end : TET3"	
BASE COUNT	201 a	64 c 131 g 202 t 503 others
ORIGIN		

Query Match	2.1%	Score 74.4	DB 17	Length 1101
Best Local Similarity	18.3%	Pred. No. 1.8e-06		
Matches 106	Conservative 241	Mismatches 233	Indels 0	Gaps 0

526 AATTCGTTACTGCTCATATTCAATTAA TTCACTATTAAATAAACAGTTCCTAAAAGGCTGT 585

Db 480 A A A T T W T T T T T T T W T T A W T T W A T A W W T T W W W T T A A A W A A A A A A A A A A A A A W A A A W A T 539

586 TTATTGGATGATATTGGAATTATCACATAATAATTGATGCTATTATTACTTGGCTAT 645

Db 540 A A A T G T G T G T G T T T Y T T W A A W A T A A A M C M A A W Y Y H T Y T T Y T T Y H Y Y T Y W T Y T M T W H Y H T M Y T 599

646 TGGTATCACTTTCATGCTCTATACATGTAATATATTTCGAGTTAGACCTTAATCAAG 705

Db 600 НАВАНТТВУННУНТГАМННМТННТМДАНННТТТТАУУУУТСТМУУННМНННАНА 659

706 TAATTTGCTATTTAATTATCTGAATAATATGTAAATCGATTGCTTTGTGGTTATTTT 765

Db 660 HAAWWTTHTWTHAUNWATYHUYUUMYSAMSMSTHTSCHNSUYUUNUTAHNTHTNHUY 719

766 TATGTTTGCATTTTAAATGACGGTGAGCTTGTGCATTCATATTTTATATGATGACAA 825

Db 720 АНУМУУМВАУУМУСТАУНУНННННУНВАУНТТУАМАНМММНННАУАААААА 779

826 CATCTTGAAGTATTTAAGATATTGTTAATGCATGAGGGTTTGGCGTATTTTAA 885

DB 780 ААТТННУННТТНМННТТМНУНУМУТССУМСТУНСМНУНТАУТСВТНННММТТНМ 839

[illegible]

RESULT	8
CNS035N7	
LOCUS	
DEFINITION	CNS035N7      576 bp      DNA      linear      GSS: 15-MAY-2000 Tetradon nigroviridis genome survey sequence PUC-Ori end of clone 21A06 of library G from Tetradon nigroviridis, genomic survey Sequence.
ACCESSION	AL228940
VERSION	AL228940.1
KEYWORDS	GI:7887933
SOURCE	GSS: genome survey sequence, Tetradon nigroviridis. Tetradon nigroviridis
ORGANISM	Tetradon nigroviridis

REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS
1 (bases 1 to 576)	Roest-Crocollus, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Benoit, A., Fizames, C., Wincker, P., Brottier, P., Quelier, F., Saurin, A. and Weissenbach, J.	Human gene number estimate provided by genome wide analysis using tetradon nigroviridis DNA sequence	Unpublished	2 (bases 1 to 576)	Roest-Crocollus, H., Jallion, O., Dasilva, C., Fizames, C., Fisher, C., Benoit, A., Fizames, C., Wincker, P., Brottier, P., Quelier, F., Saurin, A. and Weissenbach, J.

TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish <i>Tetraodon nigroviridis</i>
JOURNAL	unpublished
REFERENCE	3 (bases 1 to 576)
AUTHORS	Genoscope.
TITLE	Direct Submission
JOURNAL	Submitted (12-APR-2000)
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the <i>Tetraodon nigroviridis</i> genome. For more information, please take a look at <a href="http://www.genoscope.cns.fr/tetraodon">http://www.genoscope.cns.fr/tetraodon</a> .

FEATURES	source	location/Qualifiers
	1..576	
		/organism="Tetrahodon nigroviridis"
		/db_xref="taxon:9983"
		/clone="214A06"
		/clone_lib="G"
		/note="Genoscope sequence ID : C0AG214BA03SP1-end : pUC-Or1"
BASE COUNT	62 a	5 c 18 g 444 t 47 others
ORIGIN		

2.18; Score 74.2; DB 17; Length 576;  
Query Match Similarity 43.7%; Pred. No. 1,9e-06;  
Best Local Similarity 43.7%;  
Matches 178; Conservative 27; Mismatches 202; Indels 0; Gaps 0;

OY 542 TATTCATAATTCTCACTATTAATATAACAGTCTCTAAAGCGCTGTTTATTTGGATGAATATT 601  
          ::               :  
Db 3 TTTTTTTTTTTTTTTTTTTTGAMTTTGGATTTTGAAAAAATAAAAATCTAAATAATTAAAGAT 62  
          ::               :

[illegible]

LOCUS	DEFINITION	1101 bp	DNA	1 linear	GSS 26-JUL-1999
CNS01219/c					
CNS01219	Drosophila melanogaster genome survey sequence SP6 end of BAC BAC08024 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	REFERENCE
AL101595 AL101595.1 GSS	GI:5613206
<i>Drosophila melanogaster</i> . <i>Drosophila melanogaster</i> . Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Phytozoa; Drosophilidae; <i>Drosophila</i> . 1 (bases 1 to 1101)	

TITLE	COMMENT
<p>Direct Submission  Submitted (23-JUL-1999) Genoscope - Centre National de Séquençage :  BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr  Web : www.genoscope.cns.fr)</p>	<p>Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - <a href="http://www.edgp.ebi.ac.uk">http://www.edgp.ebi.ac.uk</a> . This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CPDH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBeloBAC11.</p>

FEATURES	source	location/qualifiers
		1. 1101
		/organism="Drosophila melanogaster"
		/db_xref="taxon:722"
		/clone="BACN08024"
		/clone_1b="DrosBAC"
		/plasmid="pBelBAC11"
		/note="end : SP6"
BASE COUNT	436 a	142 c 169 g 227 t 127 others
ORIGIN		

	Query Match	2.1%	Score 73.8;	DB 17;	Length 1101;
	Best Local Similarity	43.5%;	Pred. No. 2.5e-06;		
	Matches	178;	Conservative	38;	Mismatches 190; Indels 3; Gaps 1;
OY	543	ATTCATTATATTCACCTATTAAATAAACAAGCTGTTTATGGATGAATATTC	602		
		..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::..:::			
Db	574	ATCCCCCATTTTTTATTATTAATTCAGGGTTTTTTTAACTCTTAATTTAAATTTATTTT	- 516		
OY	603	GAAATTATCACATAAATAATTGATGCATATTATCTTGCGTATTTGGTATCAACTTTCATG	662		



Query Match 2.0%; Score 73.2; DB 17; Length 870;  
Best Local Similarity 40.6%; Pred. No. 3.3e-06;  
Matches 164; Conservative 33; Mismatches 207; Indels 0; Gaps 0;

```
OY 543 ATTCATTAATTCCTTTAATAAACAAGCTTAAAGCGCTTATTTGATGATATTC 602
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 405 ATTTAAAAATTAATAAATTAACATTTAGAACATTATATCTAGTNNNGNAANGA 346
OY 603 GAATTTATCATATATATGATGATCTATTTACTTGTCTGATGATATCACTTCATG 662
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 345 TKATNNKNNATNNNNATTTTATTTTATTTTATTTTATTTTATTTTATTTTATKT 286
OY 663 CTCTATACATGTAATATTTTGCAGTTAGACCTTAATTCAGTAATTTGCTATTAAT 722
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 285 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 226
OY 723 TATATCTGAAATATATGATATGATGCTTTGCTTATTTATTTATTTGCTTATTTT 782
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 225 ATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 166
OY 783 TAATGAGCTGAGCTTGTGATTCATATTTTATTTATGACACATCTTGTGATGATAT 842
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 165 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 106
OY 843 TTAAGATATTTGTAATGACATGAGGGTTGCGTATTTTATTTATTAATCAATATAAA 902
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 105 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 46
OY 903 ATCAACATATATTTATTTTGTCTTTTATAGTGTCTTT 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 45 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2
```

RESULT 12  
CNS0012D 991 bp DNA linear GSS 03-JUN-1999

LOCUS DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:  
BACR36013 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL074343.1 GI:4954012  
VERSION 1  
KEYWORDS GSS.  
SOURCE Drosophila melanogaster.  
ORGANISM Drosophila melanogaster.  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 991)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)

COMMENT Determiation of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazutoyo Ooeegawa and  
Aaron Mammoser in Pletier de Jong's laboratory in the Department of  
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,  
NY. The library is named RPCI-98 and was constructed by partial  
EcoRI digestion of Drosophila DNA provided by the BDGP from the  
isogenic strain y2; cn bw sp. the same strain used for the BDGP's  
P1 and EST libraries. A more detailed description of the library  
and how to order individual BAC clones, the entire library, or  
filters for hybridization from the BACPAC Resource Center can be  
found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

FEATURES  
source location/Qualifiers  
1..991  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"

BASE COUNT 176 a 95 c 64 g 515 t 141 others  
ORIGIN

Query Match 2.0%; Score 72.8; DB 17; Length 991;  
Best Local Similarity 39.0%; Pred. No. 4.2e-06;  
Matches 194; Conservative 30; Mismatches 274; Indels 0; Gaps 0;

```
OY 449 CCTCCGATTAATTTGGGTGACATATTAACCTTTGCTATCGAAGAGTTATGTTGAGG 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 57 CCTTACAGAAATTCACAAATTTAATAATTTTAAATTTCTTATTCANNNNAATKT 116
OY 509 AGCGGTTTGGTGGCTTAATTCGTTACTGTCATATTCATTAATTCATTAATTAATAA 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 117 TATATACGTGTTAKTTTWTTAATTAATKACACAAAGCTTAAATTTATCTTTTAK 176
OY 569 CAGTTCAAAAGCGCTTTATTTGATGATATTTGGAATTTATCATAAATATGATGCT 628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 177 ACCATTTTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 236
OY 629 ATTTATCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 237 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 296
OY 689 TAGACCTTAATTCAGGTAATTTGCTATTTAATTAATTAATTAATTAATTAATTAAT 748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 297 TNNNNNNNNNTNNNNNTNNNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNTNT 356
OY 749 TGCCTTGTGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 357 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 416
```

OY 809 ATTTTATGATGACACATCTTGATGAGTATTAAGATGTAATGATGACATGAGGG 868

DB 417 CTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 476

OY 869 TTTGCGTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 928

DB 477 TTTTCTCKKTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 536

OY 929 TTTTATAGTGTCTTTT 946

DB 537 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 554

RESULT 13  
CNS06PLB 1084 bp DNA linear GSS 05-JUL-2001

LOCUS DEFINITION T7 end of clone AV0AA015D02 of library AV0AA from strain CBS 379 of  
Saccharomyces exiguus, genomic survey sequence.

ACCESSION AL409509.1 GI:12177058  
VERSION 1  
KEYWORDS GSS.  
SOURCE Saccharomyces exiguus.  
ORGANISM Saccharomyces exiguus.  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 1084)  
Auteurs Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
Bolotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,  
de-Montigny,J., Dujon,B., Durieux,P., Lepingle,A., Llorente,B.,  
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potter,S.,  
Saurin,M., Tekala,F., Toffano-Nioche,C., Weslowski-Louvel,M.,  
Wincker,P. and Weissenbach,J.  
Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
yeast species for molecular evolution studies  
FEBS Lett. 487 (1), 3-12 (2000)  
JOURNAL MEDLINE  
PUBMED 20584711  
11152876  
REFERENCE 2 (bases 1 to 1084)  
Auteurs Bon,E., Neuvéglise,C., Lepingle,A., Wincker,P., Artiguenave,F.,

[illegible][illegible]



QY	944	TTT	946
Db	1137	TTT	1139

RESULT 15

LOCUS	CNS0152H	614 bp	DNA	linear	GSS 26-JUL-1999
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN12N03 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	X110416				

DB	QY	DB	QY
104	935	44	33

```
Search completed: March 1, 2003, 21:25:17
Job time : 4804 secs
```



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2003, 20:02:01 ; Search time 63 seconds  
(without alignments)  
979.286 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343

Sequence: 1 MPLSKHQIRQLSKPLSDSI.....DRGSKQDQKQSSATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq\_101002:\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152.5	6.5	1503	19	AAW48845
2	138	5.9	2008	18	AAW22016
3	136.5	5.8	7201	22	ABB71136
4	135	5.8	2013	22	AAW67964
5	134.5	5.7	1179	22	AAU37669
6	134.5	5.7	1179	22	AAW01107
7	128	5.5	2017	22	ABG06301
8	126.5	5.4	752	23	AAU84329
9	126	5.4	610	19	AAW68206
10	122	5.2	1370	22	ABB68691

11	122	5.2	1971	22	ABG07508
12	120.5	5.1	1048	22	ABW59245
13	120.5	5.1	1294	22	ABB63502
14	120	5.1	1131	22	ABG07281
15	120	5.1	1497	22	ABG20153
16	119.5	5.1	1066	22	AAW67418
17	119.5	5.1	1179	22	AAU38015
18	119	5.1	1023	23	AAW48189
19	118.5	5.1	806	21	AAW72574
20	118	5.0	1392	20	AAW06999
21	117.5	5.0	687	19	AAW41586
22	117.5	5.0	868	22	ABW59635
23	117.5	5.0	1427	12	AAW10534
24	117	5.0	2779	22	ABW62371
25	116	5.0	815	22	ABW60663
26	116	5.0	881	22	ABW20258
27	115.5	4.9	663	22	AAW94117
28	115.5	4.9	1690	22	ABW61144
29	115.5	4.9	1690	22	ABW61173
30	115.5	4.9	2519	22	ABW16636
31	114.5	4.9	955	15	AAW57365
32	114.5	4.9	955	17	AAW03691
33	114.5	4.9	1091	22	ABW68898
34	114.5	4.9	2053	22	ABW71118
35	114	4.9	952	22	AAW79258
36	114	4.9	1939	23	ABW77096
37	113.5	4.8	2542	22	ABW71337
38	113.5	4.8	2816	22	AAW68572
39	113	4.8	1411	17	AAW02258
40	112.5	4.8	733	21	AAW07449
41	112.5	4.8	1130	22	ABG12663
42	112	4.8	881	22	ABG05280
43	112	4.8	900	22	AAW20178
44	112	4.8	900	22	AAW20179
45	112	4.8	2013	22	ABB62322

#### ALIGNMENTS

RESULT 1	
AAW48845	
ID	AAW48845 standard; Protein; 1503 AA.
AC	AAW48845;
DT	12-OCT-1998 (first entry)
XX	
DE	Human receptor tyrosine kinase LMR2_h.
XX	
KW	Receptor tyrosine kinase; LMR2_h; human; signal transduction;
KW	cancer; neurodegenerative disorder.
XX	
OS	Homo sapiens.
XX	
FH	
FT	Key
FT	Peptide
FT	Domain
FT	Location/Qualifiers
FT	1..23
FT	/label= Sig_Peptide
FT	43..69
FT	/label= TMD
FT	/note= "Transmembrane domain"
FT	409
FT	/label= O-phosphorylated
FT	477
FT	/label= O-phosphorylated
FT	552
FT	/label= O-phosphorylated
FT	1032
FT	/label= O-phosphorylated
FT	1100
FT	/label= O-phosphorylated
FT	1105
FT	/label= O-phosphorylated

Novel human diago  
Drosophila melanog  
Drosophila melanog  
Novel human diago  
Novel human diago  
Novel human diago  
Amino acid sequenc  
Streptococcus pneu  
Listeria monocytog  
Human cytoskeletal  
Resilin protein seq  
Truncated resilin p  
Drosophila melanog  
Human 160kD mediat  
Drosophila melanog  
Drosophila melanog  
Novel human diago  
Human protein sequ  
Drosophila melanog  
Drosophila melanog  
Novel human diago  
K39 polypeptide of  
Leishmania chagasi  
Drosophila melanog  
Drosophila melanog  
Human protein seq  
Human alpha-myosin  
Drosophila melanog  
Human novel cytol  
Nucleolar/endosoma  
Amino acid sequenc  
Novel human diago  
Novel human diago  
S. cerevisiae mult  
S. cerevisiae mult  
Drosophila melanog

[illegible]

Oy	158	VGDSESSILYAPVLQPLVGEVTFFD-----PQSAERKG---EISQLKMLTT	203
	:	:   :	:   :
Db	523	-DDSGQDVPLRPV-----GVVFVFAHNLSVGSDYYIQLEESGSGNLEIDYPALLTT	574
Oy	204	TVAQRFLAIQRKMEAKRCVYQDLRLSALVSTKCHSIGSGOSTNGCFKSLITRENLVH	265
	:	:   :	:   :
Db	575	-----DMNPETRGPELSQLTALNSVE--LBESTIDEDFQOSTDPKDSLPG	626
Oy	264	LSGIKLAKRAEKATVEQDAVASSEVSSEGELPSHMDTKHIERIIPMASSEQA---TVSOHLH	319
	:	:   :	:   :
Db	621	DLHVTSGESEPFPNNIFNDVKSE---DLPSHQ-KTFIDLMELNGVADAFKPATLSSLD	674
Oy	320	-----AGNISLGCLNNNRDLAFHLREVSDYFPOSE	355
	:	:   :	:   :
Db	675	NPKESVITGHFEKEPRKPIFDPSEPLCLSDNIIMHDNPDLN-----VOELSENF---	723
Oy	353	PHSPISPELLKAIRMGVLSLPPELLREMSE-QNGDALSTIEFNACGLNHLD-----	404
	:	:   :	:   :
Db	724	----LFLOENLNKLGSISSKEHINDIQTLEKNNGFTPEAMLETISCNSLDELIOFAENKP	778
Oy	402	-----QVLL-----PEVSTPTVGIESPOTPOAKPSVSDPSRV	433
	:	:   :	:   :
Db	779	GLSLLQENVSTRKDQTDVWLGLDTLSLSSQSPVVQPPSPFEEIETPRRVPDPSLTQG	838
Oy	434	EEHVSQTSPPVDTSKKODOKPOSSATFSALS	462
	:	:   :	:   :
Db	839	E---TOPPCLDIVPEDECLIHODISPDAVT	864
	:	:   :	:   :

RESULT 2  
AAW22016  
AAW22016 standard; Protein; 2008 AA.  
AAW22016;  
AAW22016;  
03-OCT-1997 (first entry)  
utrophin truncated polypeptide.  
utrophin truncated polypeptide.  
utrophin; maligne; Duchenne muscular dystrophy; gene therapy.  
Synthetic.  
OS  
XX  
FH Key Location/Qualifiers  
FT Region 238..250 /note="residues 238-250 (all encoded by codon NNN) may comprise the sequence DKRSIIIMYTLTI, absolutely conserved in human, mouse and rat utrophins"  
XX  
FM W09722696-A1.  
PD 26-JUN-1997.  
XX  
PF 19-DEC-1996; 96WO-GB03156.  
XX  
PR 24-OCT-1996; 96GB-0022174.  
PR 19-DEC-1995; 95GB-0025962.  
PR 26-JUL-1996; 96GB-0015797.  
PA (MED1-) MEDICAL RES COUNCIL.  
XX  
PI Davies KE, Tinsley JM.  
XX  
DR WIJ; 1997-341687/31.  
DR N-PSDB; AAT74665.  
XX  
PT Nucleic acid encoding utrophin, truncated forms and related vectors  
PT -also transformed mammalian cells, used for alleviating symptoms of  
PT muscular dystrophy  
XX  
CS Claim 1; Fig 3; 78pp; English.

XX A truncated utrophin polypeptide (AAW22016) having utrophin function  
CC includes the actin-binding domain and dystrophin protein complex  
CC (DPC) binding domain of full-length utrophin (see also AAW22017), but  
CC lacks the rod domain. It is obtd. by expression of a utrophin  
CC minigene (AAW74665) comprising approx. the first 2 kb and the last 4  
CC kb of the full-length utrophin coding sequence (see also AAW74666).  
CC Expression of the truncated utrophin significantly decreases the  
CC severity of the dystrophic muscle phenotype in an animal model.  
CC Indicating usefulness in treatment of muscular dystrophy. The  
CC polypeptide can also be used to screen for substances that modulate  
CC utrophin binding to actin and/or the DPC.  
XX  
SQ Sequence 2008 AA:  
Query Match 5.9%; Score 138; DB 18; Length 2008;  
Best Local Similarity 21.2%; Pred. No. 0.063;  
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;  
QY 1 MPLSKHQIEQLSKPLSDSDICGVYLLKLSAFRLRNFVNAQTALRKLSQNPADERDA 60  
DB 420 MELQKQLOQLSSWLA-----LTERIQKMESEPIGLDULPS 455  
QY 61 LQEACLKWKILSDSLYEQSKTTRDIELT-----SWFVAQFLIDPTTLESANSLWLA 115  
DB 456 LQK-LLOEHSKSLQNDLEAEQKVNSLTHMVYVDENSEGSEATALLDQIQKIGRMTAVC 514  
QY 116 DLSEKHHMDLNPVLPVFTLKSDDKGEREQADAKVAFQVLDSESSILYAPVQLP 175  
DB 515 RWTERNNRL-----DEISILWQELLLEEQ 538  
QY 176 LVGEVTFEFDQSAERKEISQLKSMLTVTVAOERFAIOFKMENAKRCVTOIDRLSAL-- 232  
DB 539 CLLEAWLTERKEEALNKVQTSNFKQKELSVSRRLAILKEBMEKR--QITDQISEIGOD 596  
QY 233 -----VSTKCHSGSGSTNGFSAKSLITRYENALVHLSGIKLPAKAEAKTVEGEVA 283  
DB 597 VGQILSNPKASKKWNSSSEELTQRM--DSLQVRLEDS-----SNQVTOAVA 640  
QY 284 ESSVSEGEPLSHMDTKHIERIPMASEOQ-----TVSOH-LHAGNLSEICNLNNMRDL 336  
DB 641 KLGMSQIPQKDLLETIVAVREKGVKKPKQELPPLTKAEHAKMKRSTTELS----- 691  
QY 337 AFHLIREVSDYFROSEPH-----SPISFILLEKAIKMGYSLPELLREMMSEONGDA 387  
DB 692 --EMIQRLRLDLOEMEVHAERKVLNRTLEMLSDKS-----LSLPE--RDKISE-----S 738  
QY 388 LSTFENAGLNHLDOVLLPEVSTPTVGIESPQTFOAKPSVSDPSVEBHVSQTSPPVDQS 447  
DB 739 LRTV-----NMTWNTKICREV--PTTLKECIOEPS--SVSQTR-IAAHPVQKVLVSS 786  
QY 448 KODKPOSSATSALS 462  
DB 787 ASDIPVOSHRTSEIS 801  
RESULT 3  
ID ABB71136 standard; protein; 7201 AA.  
XX ABB71136;  
XX 26-MAR-2002 (first entry)  
DE Drosophila melanogaster polypeptide SEQ ID NO 40200.  
KW Drosophila, developmental biology; cell signalling; insecticide;  
OS pharmaceutical.  
XX Drosophila melanogaster.  
XX WO200171042-A2.  
XX

PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001MO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.  
XX  
DR N-PSDB; ABL15239.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 40200; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins  
CC (ABBS7737-ABBS72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 7201 AA:  
Query Match 5.8%; Score 136.5; DB 22; Length 7201;  
Best Local Similarity 22.0%; Pred. No. 0.54;  
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;  
QY 30 SAFPRLRNE-FNVQNTA---LRKLS-----QNSAD---ERD---ALQEA---CLKWK 70  
DB 1263 SEFYDQNDIDFSMLDTAETELRSITPLQTDPKKNVSODLSKSRDLNVOLQASHOLRLKH 1322  
QY 71 ILSDSLYEQSKTTRDIELISWFAAOFLLDTTLESANSLWLADESEKHMWD----- 124  
DB 1323 ALKSELRLAPARDKRP-LSEVTEVEKMFNTMEHKKDRGYLEDISAK-WNNYKTRLA 1380  
QY 125 -----LNPVLP--VETLSDDDKGEREQADAKVAFQVLD--SESSILYAPVQLP 175  
DB 1381 ELQEMANKVAPKNEIALQSEDLTPREHV---VKQAFKRILGDMKQDLDLAADASELAP 1437  
QY 176 LVGEVTFEFDQSAERKEISQLKSMLT----- 203  
DB 1438 KEGNIA---EAKRLKEGELTKLOEVLSAINRNDHQAQAVOEDLVNMQQFOAGLOQIKPA 1493  
QY 204 -----TVAOERFAIOFKMENAKRCVTOIDRLSALVSTKCHSGSQ 243  
DB 1494 VQSEVAVNNVNVSKRPISEAEVAMQNAQOPEPQ-----COBQDLKLG-LSNISHKMLCK 1548  
QY 244 STNFGFAKSLTRYEN--ALVHLSGILAPKAE-----KTVEGEVA--ESSVSEGE--- 291  
DB 1549 -TN---APDELDAKHSRMTAVHENAQASAKLEKLVANMKSFDAKALBDWVGQBOOM 1604  
QY 292 --LPSHMDTKHIERI-----PMASEQAQTV-----SQHLHAGLSL-- 326  
DB 1605 SRRPVLNTPHIDKLEKELVYKLSFNNEISQQAQKLTTLGOMADQISLHLAPEGAAALKD 1664  
QY 327 -----GNLNNMRDLAFHLIREVSDYF--ROSEPHSPISFILLEKAIKMGYSLPELLRE 378  
DB 1665 RVNQMKKLOKLSBATRGH--INEVSDAIIISROFNALVNF-----SNW-----MEDLN 1713  
QY 379 MMS---EONGDALST---INNAAGLNHLDO-----VLLPEVSTPTVGIESPQTFOAKPSV 427  
DB 1714 QVTOVEINPERVETSLHVIHALQEHADKKPSPAIYDEVKQALAG---ATPEESNAL 1769

```

OY 428 SDP-----RSVSEHVSQTSFV-----DPOS-----KODQKPOSSATSAL 461
Db 1770 NDATYALVYNTIONLETNMLQKKALEKWTELLGKNDTSHLNTYKHLQDKPEGPAAEEL 1829
OY 462 S 462
Db 1830 S 1830

RESULT 4
AAB67964
ID AAB67964 standard; Protein; 2013 AA.
XX
XX AAB67964;
XX AC
XX 29-JUN-2001 (first entry)
XX DT
XX
XX Amino acid sequence of utrophin B isoform minigene.
DE
XX
XX utrophin; promoter; chromosome 6q24; exon 1b; dystrophin;
KM muscle-specific transcription; muscular dystrophy.
XX
XX Unidentified.
OS
XX WO200125461-A1.
XX PN
XX 12-APR-2001.
PD
XX
XX 04-OCT-2000; 2000WO-GB03800.
PF
XX
XX 04-OCT-1999; 99GB-0023423.
PR
XX
XX (ISIS-) ISIS INNOVATION LTD.
PA
XX
XX Burton E, Tinsley J, Davies K;
PI
XX
XX WPI; 2001-273582/28.
DR
XX N-PSDB; AAF84673.

Novel nucleic acid comprising promoter for mouse, human utrophin genes,
PT for promoting tissue-specific transcription of linked sequences and in
PP screening for substances able to modulate utrophin promoter activity -
XX
XX Disclosure; Fig 9; 76pp; English.
PS
XX
XX The present sequence is encoded by an utrophin B isoform minigene.
CC The specification describes human and murine utrophin alternative
CC promoters. Utrophin is a 395 kDa protein encoded by a gene located on
CC chromosome 6q24. The alternative promoter is highly regulated, and has
CC little similarity to the synaptically expressed promoter. The
CC alternative promoter drives transcription of a widely expressed unique
CC first exon that splice into a common full length mRNA at exon 3. This
CC unique exon (called exon 1b) encodes a novel 31 peptide which may be
CC involved in binding to the muscle membrane. Utrophin alternative
CC promoters are useful for promoting tissue-specific, preferably
CC muscle-specific transcription of an operably linked sequence of
CC nucleotides. Utrophin exon 1b polynucleotides and polypeptides are
CC useful in the manufacture of medicament for treating a dystrophin
CC phenotype in a mammal. Up-regulation of utrophin expression may
CC compensate for dystrophin loss in muscular dystrophy patients.
CC
XX
XX Sequence 2013 AA;

Query Match 5 8%; Score 135; DB 22; Length 2013;
Best Local Similarity 21.2%; Pred. No. 0.11.
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19.

1 MFLSKHQIEQLSEKPLSDSDSICGVYLLKLEKSARPLRNENFVATALRKLSQSPSADERDA 60
Db 425 MFLQKKQLQQLSSWLA-----LTERKQKNESLPGLGDDPS 460
61 LQEAQILMKRKILSDSLYEFQSKTYTDIELI-----SWFYAAQFLDPTLTLESANSLIEWIA 115

```

Db	461	LQK-LLEHKSLQNDLDEQYKVASLTHMYIYDENGSESTALLEQDLGERMTAVC	5119
Qy	116	DLSEKHMDHLNPLVPLETFLKSDDDKGEREQADAKVKAFFOLVGDSESSILYAPVQLP	175
Db	520	RWTEBRMNRLL-----QETSLMQELLEEQ	543
Qy	176	LVGEVTFEFDFOESAERKEGISQLSKMLTTTVAQERFAIQFKENAKRCVYOLDRLSAL---	232
Db	544	CLLEAWLTKEEKALDKVOTSNFKQOKELSVSRRLATLKEDEMERR--QTLDOJSEIGOD	6011
Qy	233	-----VETKCHSGSOSTNGFAKSLTFRVENALVHLDSGIKLAPKAERKTVDEVA	283
Db	602	VGOLLSPNKASKKMMDSOEELTQW--DSLVRLEDS-----SNQYQAVA	645
Qy	284	ESSVSEGEPLSHMDTKHLERIPMASEQAO-----TVSOH-LHAGNLSEGNLNNMRDL	336
Db	646	KLGMSQIRQKXLLLELVHAREGAWKKPKQELRPPLTKAEHMMQKRSTTEIG-----	696
Qy	337	AFHLIREVSDYFROSEPH-----SPISFLLEKAIKMGYSLPLELREHMSSEONGDA	387
Db	697	--ENQLQELRDLTQEMEVHAERKLMNLRTLEMLSDKS---LSLPE--RDKISE---S	743
Qy	368	LSITFNAGLNLHLDQVLLPEVSTPTVGIESPQTQOAKPSVSDPSVEHVSQTSPTVDTS	447
Db	744	LRTV-----NMTWKIKREV--PTTLKECIQEPF---SVSQTR-IAHNRVQKVVLYSS	791
Qy	448	KQDKKPOSSATSALS	462
Db	792	ASDIPVQSHRTSEIS	806
RESULT 5			
AAU37669			
ID	AAU37669	standard; Protein: 1179 AA.	
xx	AAU37669;		
xx	14-FEB-2002	(first entry)	
DE	Streptococcus pneumoniae cellular proliferation protein #98.		
KW	Antisense: prokaryotic cellular proliferation protein;		
xx	antibiotic; antibacterial; drug design.		
xx	Streptococcus pneumoniae.		
PN	WO200170955-A2.		
PD	27-SEP-2001.		
PF	21-MAR-2001; 2001WO-US09180.		
PR	21-MAR-2000; 2000US-191078P.		
PR	23-MAY-2000; 2000US-206848P.		
PR	26-MAY-2000; 2000US-207727P.		
PR	23-OCT-2000; 2000US-242578P.		
PR	27-NOV-2000; 2000US-253625P.		
PR	22-DEC-2000; 2000US-257931P.		
PR	16-FEB-2001; 2001US-269308P.		
PA	(ELIT-) ELITRA PHARM INC.		
PI	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,		
PI	Yamamoto RT, Xu HH;		
DR	WPI: 2001-611495/70.		
DR	N-PSDB: AAS55528.		
xx	New polynucleotides for the identification and development of		
xx	antibiotics, comprise sequences of antisense nucleic acids -		
xx	Example 3;Seq ID No 13262; 511bp; English.		

XX The invention relates to antisense inhibitors of genes essential to  
CC prokaryotic cellular proliferation, their use in identifying the  
CC genes, their use in the discovery of novel antibiotics, the essential  
CC genes themselves and the encoded proteins. The prokaryotes used are  
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
CC invention is also useful for the identification of potential new targets  
CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins.  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequence is also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 1179 AA:

Query Match 5.7%; Score 134.5; DB 22; Length 1179;  
Best Local Similarity 21.6%; Pred. No. 0.057;  
Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

QY 5 KHOIEQLSKPLSDSICGVLYLKEKSAFRLPNEFNVAQTALRKLSQNSPADERDALQEA 64  
DB KPELEQLQKETAIDE---ASIGSEBAALKTLQDDMAALTERLEAIK---SOGEGARIQEQ 727  
QY 65 CLNWKILSDLSYEQFSKTRFDILISMFWAAGFLDPTLESANSLEMLADSEKHMHD 124  
DB 728 GLS-----LAYQTSQOYVELETL-WKLOEE-EIDRLSG-----DWQAD-KEKQES 772  
QY 125 L-----NPVLPELTAKSDDDKGRERQADAKVKAFFQVGDSESSILYAPVLOPL 176  
DB 773 LATIASKONLEAEIEIKSNKNAIQERYQ-----NLQEEVAQARILKTEL 818  
QY 177 VGEVTFDFQSAERKGEISQLKSMLTITVAQERF-AIQFKENAKRCVQDLRLSALVS 234  
DB 819 QGQR-YEVADIERLG-----KELDNLNIEQEIORMLOEKVDNLEKYDE-----LLS 866  
QY 235 TKCHLSGSOSTNF--GFAKSL--LTRVENALVHLSG-----IKLAPRAEAKTV 278  
DB 867 QQAESKTQKTNLQOGLIRKQFELDDIEGLDLDASHLDQARQONEEMIRKQTRAERK-- 924  
QY 279 EQEVAESSVSEGELPSHMDTKHIERIPMASQVQVSOHLHAGNIS--ELGNLNNMRDL 336  
DB 925 KEKVSERL-----RHQNLQDQOQISYTEALEKAHLEMLNLADQEVQDDEKAIKRL 977  
QY 337 AFHLRLVSDYFRQSEPHSPISLEKAIKRGYISLPPELLREMSSEONGDA--LSTTFN 393  
DB 978 GPNVLEAIDY---EEVHNRLDPL--NSQRDILISAKNLLETITENMDVEKREKSTFE 1032  
QY 394 AA-----GLNHLDOVLLEPVESTPTVGIESPOTPOAK 424  
DB 1033 AIRESEKVTFRKMGCGQAD-LILTEGDLTAGVEISVQPPGK 1074

RESULT 6  
AAM01107  
ID AAM01107 standard; Protein; 1179 AA.  
XX

AC AAM01107;

XX 02-OCT-2001 (first entry)

XX CFE 110 protein sequence.

XX Antibacterial; vaccine; gene therapy; bacterial cell wall viability;  
KW CFE; Conserved Essential Gene; bacterial infection;  
KW antisense therapy; antibiotic resistance.

XX Streptococcus pneumoniae.  
OS  
XX  
XX WO200149721-A2.  
PN  
XX  
XX 12-JUL-2001.  
PD  
XX  
XX 29-DEC-2000; 2000WO-US35604.  
PF  
XX  
XX 30-DEC-1999; 99US-0174089.  
PR  
XX  
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.  
PA  
XX  
XX Dougherty TJ, Pucci MJ, Dougherty BA, Davison DB, Brucoleri RE;  
PI Thannasi JA;  
PI  
XX  
XX WPI; 2001-496721/54.  
DR  
XX  
XX N-PSDB; AAH90806.  
DR  
XX  
XX Nucleic acids encoding conserved essential genes involved in bacterial  
PT replication which are potential targets for the treatment of antibiotic  
PT resistant bacterial infections -  
PT  
XX  
XX  
XX Claim 27; Pages 365-369; 380pp; English.

CC The present invention relates to nucleic acids (AAH90701-AAH90918)  
CC encoding polypeptides (AAM01002-AAM01114), which are essential for the  
CC viability of a bacterial cell wall. The acronym CFE stands for "CEG For  
CC Expression", where CEG stands for "Conserved Essential Gene". The nucleic  
CC acids are useful for detecting the presence of proteins essential for the  
CC viability of a bacterial cell wall in samples such as cells, tissues,  
CC biological fluids, blood, serum, nose, ear or throat swabs with ligands,  
CC and for detecting corresponding target nucleic acid molecules with  
CC complementary sequences. The nucleic acids are also useful for  
CC determining whether a genomic nucleotide sequence of interest is  
CC essential for viability of a bacterial cell or whether it resides within  
CC an operon, by integrating an exogenous nucleotide sequence comprising a  
CC portion of an open reading frame of the genomic sequence of interest  
CC (comprising 200-500 base pairs) into the genomic sequence of interest  
CC which confers a selectable phenotype to the cell, and determining cell  
CC viability with a selection agent such as chloramphenicol. The nucleic  
CC acids and proteins are also useful as vaccines and for treating bacterial  
CC infections with gene therapy and antisense therapy. The nucleic acids  
CC also enable identification of targets suitable for the treatment of  
CC antibiotic resistant bacterial infections.

XX Sequence 1179 AA:

Query Match 5.7%; Score 134.5; DB 22; Length 1179;  
Best Local Similarity 21.4%; Pred. No. 0.057;  
Matches 99; Conservative 80; Mismatches 180; Indels 103; Gaps 23;

QY 5 KHOIEQLSKPLSDSICGVLYLKEKSAFRLPNEFNVAQTALRKLSQNSPADERDALQEA 64  
DB 674 KPELEQLQKETAIDE---ASIGSEBAALKTLQDDMAALTERLEAIK---SOGEGARIQEQ 727  
QY 65 CLNWKILSDLSYEQFSKTRFDILISMFWAAGFLDPTLESANSLEMLAD-----L 117  
DB 728 GLS-----LAYQTSQOYVELETL-WKLOEE-EIDRLSG-----DWQADKEKQESL 773  
QY 118 SEKHMDLNPVLPETLKSDDDKGRERQADAKVKAFFQVGDSESSILYAPVLOPLV 177  
DB 774 ATIASDKONLEAEIEIKSNKNAIQERYQ-----NLQEEVAQARILKTELQ 819  
QY 178 GEVTFDFQSAERKGEISQLKSMLTITVAQERF-AIQFKENAKRCVQDLRLSALVST 235  
DB 820 GQKR-YEVADIERLG-----KELDNLNIEQEIORMLOEKVDNLEKYDE-----LLSQ 867  
QY 236 KCHLSGSOSTNF--GFAKSL--LTRVENALVHLSG-----IKLAPRAEAKTV 279  
DB 868 QAESKTQKTNLQOGLIRKQFELDDIEGLDLDASHLDQARQONEEMIRKQTRAERK--K 925  
QY 280 QEVAESSVSEGELPSHMDTKHIERIPMASQVQVSOHLHAGNIS--ELGNLNNMRDLA 337

Db 926 EKVSRL-----RHLONLTDQYQISYTEALEKAHELENNILAEOVODLEKAIKRLSG 978  
OY 338 FHLLREVSDFRQSEPHSPISFLLEKAIKRWGLSLPELLREMSQONDA---LSTINA 394  
Db 979 PVNLEAIQY---EEVHNRRLPL--NSQRDILSKNLLLELTTEMNDEKEREKSTEEA 1033  
OY 395 A-----GLNHLDOVLLPEVSTPTVGIESPQTPOAK 424  
Db 1034 IRESFKVTFKQMGGGGQAD-LILFEGDLITAGVEISVQPPCK 1074

RESULT 7  
ABG06301  
ID ABG06301 standard; Protein: 2017 AA.  
XX  
AC ABG06301;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #6292.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
PA (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
DR N-PSDB: AAS70488.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 20; SEQ ID No 36660; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIGO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 2017 AA;  
XX  
XX

Query Match 5.5%; Score 128; DB 22; Length 2017;  
Best Local Similarity 21.0%; Pred. No. 0.43;  
Matches 126; Conservative 77; Mismatches 191; Indels 206; Gaps 26;  
OY 4 SKHQIEQLSKPLSDSICGVYKLT-----EKS-----AFRLRNEFVNAQ 43  
Db 465 AENEIMRLSSINQNSILAEDNLIKLMRIEVELEKESLSIOKELOSLILNNEYEVIK 524  
OY 44 -TALRKLS-----QNSADERDALQEA 64  
Db 525 STATPDISLDELHDRLNLEAKKEQELNQSISEKETLIAETIEEDRQOEAFKIMILKA 584  
OY 65 CLNKKILSDSLYEQFSKTPRDIEP-----ISMVAAQFLDPTLTESANSU- 111  
Db 585 QLSKOQNEGDSIIISKLODLNDEKKRYHQLEDKMDIKTELDVQFVLLIOSEVALNDLHL 644  
OY 112 --EWLADLSEKHWDLNLFVLETLSDDDKGKEREQADAKVAFVOLVGSSESSIIYA 169  
Db 645 TKQKLEDKVENLVQDLN-----KQSESNVSIQKEKLEKEHRIQ-----NEEELSRIRN 693  
OY 170 PVLO-LPLVGEVTPFPDQSAERKGEISOLKSMLT-----TTVAQRRFAIQFMEAN- 220  
Db 694 ELMOSLNODSNSNEKDTLLKEREAEVRNLKONLSELOLNENLKKVAFDVAMENEKLVLA 753  
OY 221 -----RCVTQDLRLS-----ALVSTKCHSLGSQSTNFGFAKSLT-----TRVENALVH 263  
Db 754 CEDVRAHQLEBCLAGNNOISLEKNITVEITLKMKEGIEALELQWAKKRILLEANKYEKTEE 813  
OY 264 LSG-----IKLAPRAEAKTYEQEVAESSVSEGBELPSHMDTKHIERIPMA 307  
Db 814 LSMARNLNTSALOIEHEHLIKLNQK-----DMEIAELKKN-----IEQMDHRETKDVL 864  
OY 308 SEQAQTVSOHLHAGNLSEGLNMMNNRDLAFHLLREVSDYROSPPHSPISFLEK---A 364  
Db 865 SSSLEEQKQ-----LTQLIN-----KKEITFEKLEKRSKIQEE-----LDKTSQA 905  
OY 365 IRWGYLSPELLRE-----MSPONGDALSTIFNAGLNHLDOVL--LPEVSTPT 412  
Db 906 LRKN-----ETLRQITEEKDRSLSGMKEN-----NHLQEBELRLREBQRT 947  
OY 413 VGIESPQT---PQAKPSVSPRSVVEEHV-----SOTSVDVTQSKODQKQ 454  
Db 948 APVADPKTLDVTELAISEVSQILNTIKEHLEEEIRKHOKIIEDQNSKQQLQSLQEOKE 1007

RESULT 8  
AAU84329  
ID AAU84329 standard; Protein: 752 AA.  
XX  
XX AAU84329;  
XX  
XX 08-MAY-2002 (first entry)  
XX  
XX Protein DCC1 differentially expressed in breast cancer tissue.  
XX  
XX Human: diagnosis of breast cancer; endometrial cancer; breast tumour;  
KW MAI: mitotic activity index; cytostatic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200210436-A2.  
PD 07-FEB-2002.  
XX  
XX 27-JUL-2001; 2001WO-US23642.  
PF  
XX  
XX 28-JUL-2000; 2000US-222093P.  
PR  
XX  
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL INC.  
PA (BAK/) BAK J.  
XX  
XX Baak J, Mutter GL;  
PI







PT responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XS  
PS  
XX Claim 20: SEQ ID NO 37867; 103bp; English.  
CC  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1971 AA;

Query Match	5.28	Score 122	DB 22	Length 1971
Best Local Similarity	19.58	Pred. No. 1.3		
Matches 103	Conservative 95	Mismatches 210	Indels 120	Gaps 25

```

OY 14 PLSDSDICGVYLK--LEKSAFRLLENENNAVOTALRKU-----49
Db 239 PATSSVWGCTOMNNGISAFVGCNE--DLEVELVLSLEEDLPYRLKVTPLFYEQDM 297
OY 50 SONSADER--DALOEACLN---KWKLLSDLSLEQFSKTRTDIELLSFWAAOFLDPTL 104
Db 298 TOSPHQORHASDALPYLAEBETFRYMLIGTDYVOMKRTYNDIDMVIHLAER---DRDL 354
OY 105 ESAANSLEMLD-----LSEKH-----WDLNVLVLPETLKSDD-----DKGERE 145
Db 355 ELARIGALLKRNHVLSEQNESLEEDOLGOAFDVOQ--LOHELCKKELLRIYIASSES 413
OY 146 QADAKVAFOLVDSEBSSTILVPIQLPVGE--VTFEDFQSAERKEISOLKSMULTT 204
Db 414 ETDSSCSYPLRF---NESFSLSGGLQLEMLQELKELEENMALRSKACHITVEVTV 469
OY 205 VAQEFATQFMENAKRCVYDOLDLSALVSK-----CHLSGOSTNGFA 250
Db 470 EEKEQOLVSDCVKELRETNAMQSMATELSKSDDXDTYKSPFLCHRLKLTFFHKLGIT 529
OY 251 KSLLTRVENALVHLGKILAPKAEKTVQEOVAESSVSEGLPSHMDTKH---IERIPMA 307
Db 530 KEOLU--GKVLPH-----LPLSTENNLNGENSEIITK--EMLRLRESEHKTKLEQJHM 581
OY 308 SEQAOQ--VSQHLHAGNLSLSEGLNLNNNRDLAF--HL---LREVSDFROSEPHSPISFL 361
Db 562 QEOQKSLDIGNOMNVSEETKVTNIGNOOIDVFSHIGADLLTGSDSENKED-----632
OY 362 EKAIRMGVLSPELLREKMSBONDALSTIFNAAGLHMLDQVLLPEVSTPVGIESPO--419
Db 633 -----GALNVPYPGAKPTQGRPD-----MSALNNLFGOPKPKVSMNQLSQOKRNM 679
OY 420 -----TPQAKPSV--SDRPSVEHNSQSPVUDTQSKQOQKPOSSATSLAL 461
Db 660 LNOEVPVPSATGSSVWGCTOMNNGISAFVGCNEB--DLEVELVLSLL 726

```

RESULT 12  
ABB59245  
ID ABB59245 standard; Protein; 1048 AA

XX ABB59245;  
AC  
XX  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 4527.  
XX  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
XX  
XX  
XX  
OS Drosophila melanogaster.  
XX  
XX  
PN WO200171042-A2.  
XX  
XX  
PD 27-SEP-2001.  
XX  
XX  
PP 23-MAR-2001; 2001WO-US09231.  
XX  
XX  
PR 23-MAR-2000; 2000US-191637P.  
XX  
PR 11-JUL-2000; 2000US-0614150.  
XX  
XX  
PA (PEKE ) PE CORP NY.  
XX  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX  
DR WI: 2001-656860/75.  
XX  
XX  
DR N-PSDB; ABL03348.  
XX  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
XX  
XX  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX  
XX  
PT interactions -  
XX  
XX  
PS  
PS Disclosure: SEQ ID NO 4527; 21pp + Sequence Listing; English.  
XX  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX  
XX useful in developmental biology and in elucidating cell signalling and  
XX  
XX cell-cell interactions in higher eukaryotes for the development of  
XX  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX  
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
XX  
XX sequences (ABL01840-ABL16175) and the encoded proteins  
XX  
XX (ABB57737-ABB72072).  
XX  
XX The sequence data for this patent did not form part of the printed  
XX  
XX specification, but was obtained in electronic format directly from WIPO  
XX  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

PS Disclosure; SEQ ID NO 4527; 21pp + Sequence Listing; English.

[illegible]



CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 1131 AA:

Query Match 5.1%; Score 120; DB 22; Length 1131;  
Best Local Similarity 22.2%; Pred. No. 0.86;  
Matches 77; Conservative 60; Mismatches 122; Indels 88; Gaps 18;

QY 8 IEQLS--KPLSDSICGVYKLEKSAFRLPNEFNVAOTLKRISQNSADERDALQDAC 65  
DB 368 VKELSFAPMSDRSGDISREIDVASCRIKGSYR---ALPKYQOPKSGRFAICREV 423  
QY 66 LN-----KWKILSDLYQFSKTTTIDELISFVAQFLDITLTSANSLSEWLADLS 118  
DB 424 LNDTWSPFMS--EDSTFVSSKTPYE--BQLHRCDEDEFLDVLTLNATITLVSVQ 480  
QY 119 EKHMHLNPLVP--VETLKSDDDKGEREQADAKVAFOLVGDSE--SILYAPVLQ 173  
DB 481 KK---LSRNAPEDQEKFRLLDLSGTSVYQR--RAIYRIGYQKAPETIESLKKNPVTA 534  
QY 174 LPLV-----GEVTFPFOQAERK---GEISOLKSMLTYYAQRFAIOFKMENAKRC 222  
DB 535 VPVVLKRLKAKEEMREAOOGFNKIMREOYEKAYLKL-----DQAVNFKQNDYR-- 585  
QY 223 VTQLDRLALVSTKCHSLGSGSTNFGAKSILTFVENAL-----VHLSGIKLAPKAERKT 277  
DB 586 -----ALRS-----KSLNLETESVYDEHQBQHSBGASAPPSAHL 621  
QY 278 V---EQEVAESSYSEGLPSHMDTKHIERIPMASEQAQ--TVSQHLH 319  
DB 622 IFVYEDRQILEDA---LISYVVKRQPAIQKEDQGTIQLHLH 661

RESULT 15  
ABG20153  
ID ABG20153 standard; Protein: 1497 AA.  
XX  
AC ABG20153:

DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #20144.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.

PN WO200175067-A2.

XX 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
PA (HYSE-) HYSEQ INC.  
PI Dermanac RT, Liu C, Tang YT;  
DR WPI; 2001-639362/73.  
XX N-PSDB; AAS84340.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.

Claim 20; SEQ ID NO 50512; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 1497 AA:

Query Match 5.1%; Score 120; DB 22; Length 1497;  
Best Local Similarity 21.9%; Pred. No. 1.3;  
Matches 114; Conservative 73; Mismatches 181; Indels 152; Gaps 27;

QY 26 KLEKSAFRLPNEFN--VAQTALRKLSONPSADERDALQELCKWKILSDLYQFSKTT 84  
DB 493 QLRHDTFSSVROELEAVAGAVLSSPGSPGAVGAEOQT-----ALLBQYSEL 540  
QY 85 RDIELISFVAQAQFLDITLTSANSLSEWLADL-----SEKHMHLNPLVPETL 134  
DB 541 REND--DLFWMPRIVNTSLATEGG---LVDMGSKYRPREVPDSKPSHLK---DTV 589  
QY 135 KSDS---DKGK--ERDQADAKV---KAPFOLVGDEESSILYAPVLQ---LP--LV 177  
DB 590 RGEPNLSLEDKGRISRGNRGQRTVLGPTQVFLANKDSGYOIYVNSMKAQOEFLPKKIS 649  
QY 178 GEVTFPFOQAERKEISOLKSMLTYYAQRFAIOFKMENAK-----RCVTQDRL 229  
DB 650 GDMRGIOYKWKESSESGEERVKS-----KDSFHKLMKMDKOSTIEMELRKVTSATIEE 702  
QY 230 SALVSTKCHSLGSGSTNFGAKSILTFVEN-----ALVHLSGIKLAPKAERKVBQVA 283  
DB 703 AALDSSE-----LLTNMEDDDDTDTLTSLN-----EIAFLNQLN 739  
QY 284 ESSVSEGLPSHMDTK-----HIERIPMASEQAQTVSOHLHAGNLSLGNLNNNRD 335  
DB 740 DDSVGLAELFSSMDTEPPGARAFISKVPPGSNATQV--EHLGTC--LKELPDVQGSDDS 797  
QY 336 LAFHLREVSDYFROSEPHSPISFLLEKAIKRWGYLSPELLREMSBONDALSTIFNAA 395  
DB 798 ISPLLHLLEDDDDFSENEKQ-----LAEPN-----SEPDLKVNIDSEIKDFL--LSNKK 844

OY 396 GLNHLDOVLLPEV-----STPTVGIESPQTPQAKP-----SYSDPRS 432  
Db 845 SYGMRKEYFMPCCRARIGFSVRDAMPSTQOGLTPAEFPNRSRRRRSCRCQCSRWPRS 904  
OY 433 -----VEEHVSQTSPPVDTSKODOKPOSSATSALSW 463  
Db 905 PTFIPTSCESRSRSH-NMSSPOELQ-KPNRAMASSNRALRSW 942

Search completed: March 2, 2003, 03:24:50  
Job time : 77 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 02:21:48 ; Search time 24 Seconds  
(without alignments)  
567.617 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343  
Sequence: 1 MPLSKHQIEQLSKPLSDSI.....DTOSKODKPOSSATSLSW 463

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	152.5	6.5	1503	US-08-976-255-14	Sequence 14, Appl
2	126	5.4	610	US-09-336-447A-11	Sequence 11, Appl
3	120.5	5.1	999	US-08-770-301A-1	Sequence 1, Appl
4	120.5	5.1	999	US-09-175-581-1	Sequence 1, Appl
5	119.5	5.1	1066	US-09-541-782-8	Sequence 8, Appl
6	119.5	5.1	1066	US-09-723-820-8	Sequence 8, Appl
7	114.5	4.9	955	US-08-006-676B-1	Sequence 1, Appl
8	114.5	4.9	955	US-08-282-845-2	Sequence 1, Appl
9	114.5	4.9	955	US-08-428-414A-3	Sequence 1, Appl
10	114.5	4.9	955	US-08-428-414A-3	Sequence 1, Appl
11	114	4.9	1939	US-09-310-187A-1	Sequence 1, Appl
12	110.5	4.7	675	US-08-317-522A-9	Sequence 9, Appl
13	110.5	4.7	675	US-08-439-818A-9	Sequence 9, Appl
14	110.5	4.7	675	US-08-751-965-9	Sequence 9, Appl
15	110.5	4.7	675	US-08-738-975-9	Sequence 9, Appl
16	110.5	4.7	675	US-08-728-626-9	Sequence 9, Appl
17	110.5	4.7	675	US-08-808-599A-9	Sequence 9, Appl
18	106.5	4.5	715	US-08-849-212-6	Sequence 6, Appl
19	106.5	4.5	1234	US-09-592-054-8	Sequence 8, Appl
20	106.5	4.5	2409	US-09-592-054-8	Sequence 8, Appl
21	106	4.5	1130	US-08-519-547A-6	Sequence 6, Appl
22	106	4.5	1481	US-08-616-844-40	Sequence 40, Appl
23	106	4.5	1481	US-08-599-654-40	Sequence 40, Appl
24	106	4.5	1481	US-08-944-868A-40	Sequence 40, Appl
25	106	4.5	1481	US-08-944-868A-40	Sequence 40, Appl
26	106	4.5	1481	US-08-944-868A-40	Sequence 40, Appl
27	105.5	4.5	1312	US-08-687-080-51	Sequence 51, Appl

28	105.5	4.5	1886	US-08-938-105-3	Sequence 3, Appl
29	105.5	4.5	3696	US-09-134-001C-5080	Sequence 5080, Ap
30	105	4.5	2285	US-09-308-375-2	Sequence 2, Appl
31	104.5	4.5	534	US-09-103-664A-2	Sequence 2, Appl
32	104.5	4.5	756	US-09-085-199B-9	Sequence 9, Appl
33	104.5	4.5	1312	US-08-592-126-148	Sequence 148, App
34	104.5	4.5	2137	US-09-134-001C-4463	Sequence 4463, Ap
35	104	4.4	2101	US-08-466-390-4	Sequence 4, Appl
36	104	4.4	2101	US-08-470-950-4	Sequence 4, Appl
37	104	4.4	2101	US-08-467-781-4	Sequence 4, Appl
38	104	4.4	2101	US-08-195-487-4	Sequence 4, Appl
39	104	4.4	2101	US-08-483-924-4	Sequence 4, Appl
40	104	4.4	2101	US-09-452-294-1	Sequence 1, Appl
41	104	4.4	2101	US-09-452-294-1	Sequence 1, Appl
42	104	4.4	2293	US-09-368-590-2	Sequence 2, Appl
43	104	4.4	2482	US-08-328-254-6	Sequence 6, Appl
44	104	4.4	3248	US-08-353-700-1	Sequence 1, Appl
45	104	4.4	3248	US-08-353-700-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-08-976-255-14  
; Sequence 14, Application US/08976255  
; Patent No. 6136581  
; GENERAL INFORMATION:  
; APPLICANT: Jono, Keith E.  
; APPLICANT: Plovman, Gregory  
; TITLE OF INVENTION: KINASE GENES AND USES  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB  
; MEDIUM TYPE: storage  
; COMPUTER: IBM compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FASTSEQ for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/976,255  
; FILING DATE: No. 6136581ember 21, 1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/031,675  
; FILING DATE: No. 6136581ember 22, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1503 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; US-08-976-255-14  
Query Match 6.5%; Score 152.5; DB 4; Length 1503;  
Best Local Similarity 21.4%; Pred. No. 46-05;  
Matches 122; Conservative 75; Mismatches 189; Indels 183; Gaps 25;

```

QY 3 LSKHQIOLSPLEBD---DSICGYLLKLEKSAFPLENNENVAQTALRKSJOMNSAPERD 59
Db 370 LPRQLEO---PYDKRYTEVYLQFOWLSPEK---RPAEDVHRLILTYLRLOSQRDS--EVD 421
QY 60 ALQEAQCNKRWKILSDSYLEQFSKTRTDIELISWEVAQF-----LDDTLLESANSLEW 113
Db 422 FEQO-----WNAALPKNNINSRDSNNAAPILHDHAPARDRLGMEVEVLTVIETSGLSFEY 476
QY 114 LADLSE-KHMD-----HLN-----PYLPVETLKSDDDKKEQADAKYKAFQOL 157
Db 477 VWEAKHHDHPEDRSGHLDGLSYTSLFYVEVEFESSLSDPGPKQ-----522
QY 158 VGDSSEESILVAPYLQPLVGEVTFPD-----FQSAEKG---EISQLKSMLT 203
Db 523 -DBSGQDYPLKVP-----GVYVYFAHNLSVCSDIYTQLEEKSGSNLEIDPPLALLT 574
QY 204 TVAQERFATQKMEKNAKCVQTDRLSALVSTKCHSLGSOSTNGFASKSLTYREVALVH 263
Db 575 -----DMONPETGPELSQLALRSE---LEESTDEDFQSSSTDPRDSSLPG 620
QY 264 LSGIKLAPKAAKVEQOEVAEVAESSVSEGBELPSHMDTKIIRIPMASEOQ---TVSOHLH 319
Db 621 DLHATSGESEPNNIIFNDVKSE---DLPSHQ--KIFDLMELNGVADKPAVLSSSLD 674
QY 320 -----AGNSELGJNNMMNRDLAFHLIRVSYFQOSE 352
Db 675 NPKESVITGHPEKEKPRKIFDSEPLCLSDNNMODNDPDLN-----VOBLSBNF---723
QY 353 PHSPIFLLEKAIKRWGYLSPELLEKREMSK-QONDAALSTIFNAAGLNHLD-----401
Db 724 -----LFGQEKNNLLKGLSSKSEKHINDQOTELKNNGFTFEMLETSRNSLDELOFAENKP 778
QY 402 -----OYLL-----PEVSTPYVIGIESPQTPQAKPSVSDPSV 433
Db 779 GLSLDQENVSTKGDQTDVMTGLDGLSTLSSQSPPEVQVPTSFTEEMPRRVPDPSPLOG 838
QY 434 EHHVSQTSPPVDTOSKQOQKPPSSATSLAS 462
Db 839 E---TQPTCLDVIVPEQDCLHDDISPDVAT 864

```

```

RESULT 2
US-09-336-447A-11
: Sequence 11, Application US/09336447A
: Patent No. 6310190
: GENERAL INFORMATION:
: APPLICANT: HANSEN, ERIC J.
: APPLICANT: AEBI, CHRISTOPH
: APPLICANT: COPE, LESLIE D.
: APPLICANT: MACIVER, ISOBEL
: APPLICANT: FISKE, MICHAEL J.
: APPLICANT: FREDENBURG, ROSS A.
: TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS
: FILE REFERENCE: AMCY.024
: CURRENT APPLICATION NUMBER: US/09/336,447A
: CURRENT FILING DATE: 1999-06-21
: NUMBER OF SEQ ID NOS: 98
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 610
: TYPE: PRT
: ORGANISM: Moraxella catarrhalis
US-09-336-447A-11

```

Query Match	5.48;	Score 126;	DB 4;	Length 610;
Best Local Similarity	20.88;	Pred. No. 0.0029;		
Matches	82;	Conservative	71;	Mismatches 166;
			Indels	76;
			Gaps	15;

```

OY      43 QATALRISQNSPADERDALQ--ECLNKKMIL-----SLSYEQFSKITRDEL--- 89
      | :: || | : : : : | : | | |
DB      37 QDSISKLQ---DDIDLKQDQKMKYLLLNQLANTLITDELNNVTKNTSLEALGD 92

```

```

0Y  90 -ISMVAAQFLDPTTSSAASLSEWMLADSKSHMDHLNPLVPETLKSDDCKEKEQAD 148
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  93 EIGWLENDIADLEGVBEELTKNQMTLLEKDEH-----DRLAQNQAD 135

0Y  149 AK-----VKAFFOLVGB--SEESSIIYAPVLOJPLVEYTFPFDFOSAERKEISOLKSN 200
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  136 IOTLENNVVELLFNLSGLIDQEDAI----AKNNA5EELVDPDNEVAERIGEIHAYTEE 191

0Y  201 LTTTVAQERFAIÖFRMENAKCVTÖDLRLSALVSTKCHSL-----GSOSTNFGAKSLYLR 256
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  192 VNKT-----LENLITNSVKNNTINIKRKADIDNNINHIYELAAOODDQSHSDIKTLKNN 244

0Y  257 VBNALVHISGCIKAPKAFAKTVBEVBA5SSVSEBELPSHMDKHIERTIPASQOQTVSQ 316
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  245 VEEGLELISGHIIDÖKADL--TKDIKALESNVEEGL--DL5GRLLDOKADLYLDIKALES 301

0Y  317 HLHAGNLSGELNUNMNRDLA-----PHLREVSDFYRÖSEPHSPISFLEKRIWMC 368
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  302 NVEBGLDLSGRLLDÖKADIAQONOTIDDLAAYNELODQYÖQKOTEIDA--LNKASSEN 359

0Y  369 YLSPEL-----LREMMSEONGDALSTIFNAAGLN 398
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  360 TÖNIEDLAAYNELODATAKQÖTEIDALNKASSSN 394

```

```

1      RESULT 3
2      US-08-770-301A-1
3      : Sequence 1, Application US/0870301A
4      : Patent No. 5948637
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: IKEDA, JUN
9      : APPLICANT: KANEDA, SUDOIKO
10     : APPLICANT: YANAGI, HIDEKI
11     : APPLICANT: MATSUMOTO, MASAYASU
12     : APPLICANT: YURA, TAKASHI
13     : TITLE OF INVENTION: NOVEL STRESS PROTEINS
14     : NUMBER OF SEQUENCES: 12
15     :
16     : CORRESPONDENCE ADDRESS:
17     :
18     : ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
19     : STREET: PO BOX 747
20     : CITY: FALLS CHURCH
21     : STATE: VA
22     : COUNTRY: USA
23     : ZIP: 22040-0747
24     :
25     : COMPUTER READABLE FORM:
26     : MEDIUM TYPE: Floppy disk
27     : COMPUTER: IBM PC compatible
28     : OPERATING SYSTEM: PC-DOS/MS-DOS
29     : SOFTWARE: PatentIn Release #1.0, Version #1.30
30     : CURRENT APPLICATION DATA:
31     : APPLICATION NUMBER: US/08/770,301A
32     : FILING DATE: 20-DEC-1996
33     : CLASSIFICATION: 435
34     : ATTORNEY/AGENT INFORMATION: 435
35     : NAME: MURPHY JR, GERALD M
36     : REGISTRATION NUMBER: 28,977
37     : REFERENCE/DOCKET NUMBER: 1422-287
38     : TELECOMMUNICATION INFORMATION:
39     : TELEPHONE: (703) 205-8000
40     : TELEFAX: (703)-205-8050
41     : INFORMATION FOR SEQ ID NO: 1:
42     : SEQUENCE CHARACTERISTICS:
43     : LENGTH: 999 amino acids
44     : TYPE: amino acid
45     : TOPOLOGY: linear
46     : MOLECULE TYPE: peptide
47     :
48     : US-08-770-301A-1

```

Query Match 5.18; Score 120.5; DB 2; Length 999;

Best Local Similarity 22.28; Pred. No. 0.021;  
Matches 89; Conservative 63; Mismatches 122; Indels 127; Gaps 19;

OY 81 SKTRDIELISWVAAQFLDDTLESANSEWLADLSEKHDHLPVLPVETLKSDDDK 140

```
Db 690 ARRRRVEEIGVELVLDLPDLPEDKLAOSVOKLOD-----TLR--DLE 732
Qy 141 GKREQADAKVKAFFOLVDSESSILVAPVLQPLVGEVTFPFDPSAERKGIISOLKSM 200
Db 733 KQERKAANSLEAFI-----FETQDKLYOPEYO-----EVS--TEQREIEISGLISAASVW 781
Qy 201 L-----TTVAQERFA-----IOPKMERAKRCVQLDRLSLALVSTKCHSLGSQST 245
Db 782 LEDEGVGATTVMLEKELAEIRKLCOGLFPRVERKKWP---EKLSDLNLNHS----- 832
Qy 246 NFGFAKSLTTRVENALVHLSGIKLAPKAA--KTYQOEVAESSVSSEGLPSHMDTKHIER 303
Db 833 -----SMFLKGARLIPEMDQIFTEVEMTLLEKVINETWAMKNTATLAEQAK 877
Qy 304 IPMASQAOQTVSOHLHAGNLSELGNLNNNRDLAFLHLEVSDYFRQSEPHSIFSLER 363
Db 878 LP-ATEKPVLLSKDIEAKMA-----LDREV-----QYLNK 908
Qy 364 AIRMGYLSLPPELLREKMSQNGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTPOA 423
Db 909 A-----KFTKP---RPRPKKNGTRAEPLNANASDQGEKVIIP-----AGQTEDA 951
Qy 424 KPSVSDPSRVEEHVSQTSVPDQO-----SKODKPOSS 456
Db 952 EP-ISEPEKVE---TGSEPGDTPEPLDGGPAGAEPEQKEQST 988
```

## RESULT 4

US-09-175-581-1  
Sequence 1, Application us/09175581  
Patent No. 6034232

## GENERAL INFORMATION:

APPLICANT: IKEDA, JUN  
APPLICANT: KANEDA, SUMIKO  
APPLICANT: YANAGI, HIDEKI  
APPLICANT: MATSUMOTO, MASAYASU  
APPLICANT: YURA, TAKASHI  
TITLE OF INVENTION: NOVEL STRESS PROTEINS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
STREET: PO BOX 747  
CITY: FALLS CHURCH  
STATE: VA  
COUNTRY: USA  
ZIP: 22040-0747  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/175,581  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/770,301  
FILING DATE: 20-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR, GERALD M  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1422-287  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)-205-8000  
TELEFAX: (703)-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 999 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-175-581-1

Query Match 5.1%; Score 120.5; DB 3; Length 999;  
Best Local Similarity 22.2%; Pred. No. 0.021;  
Matches 89; Conservative 63; Mismatches 122; Indels 127; Gaps 19;

```
Qy 81 SKTRDIEILSMFVAQFLDPTLESANSLLEMLADLSEKHMDHLPVLPVETLKSDDK 140
Db 690 ARRRRVEEIGVELVLDLPDLPEDKLAOSVOKLOD-----TLR--DLE 732
Qy 141 GKREQADAKVKAFFOLVDSESSILVAPVLQPLVGEVTFPFDPSAERKGIISOLKSM 200
Db 733 KQERKAANSLEAFI-----FETQDKLYOPEYO-----EVS--TEQREIEISGLISAASVW 781
Qy 201 L-----TTVAQERFA-----IOPKMERAKRCVQLDRLSLALVSTKCHSLGSQST 245
Db 782 LEDEGVGATTVMLEKELAEIRKLCOGLFPRVERKKWP---EKLSDLNLNHS----- 832
Qy 246 NFGFAKSLTTRVENALVHLSGIKLAPKAA--KTYQOEVAESSVSSEGLPSHMDTKHIER 303
Db 833 -----SMFLKGARLIPEMDQIFTEVEMTLLEKVINETWAMKNTATLAEQAK 877
Qy 304 IPMASQAOQTVSOHLHAGNLSELGNLNNNRDLAFLHLEVSDYFRQSEPHSIFSLER 363
Db 878 LP-ATEKPVLLSKDIEAKMA-----LDREV-----QYLNK 908
Qy 364 AIRMGYLSLPPELLREKMSQNGDALSTIFNAGLNHLDOVLLPEVSTPTVGIESPOTPOA 423
Db 909 A-----KFTKP---RPRPKKNGTRAEPLNANASDQGEKVIIP-----AGQTEDA 951
Qy 424 KPSVSDPSRVEEHVSQTSVPDQO-----SKODKPOSS 456
Db 952 EP-ISEPEKVE---TGSEPGDTPEPLDGGPAGAEPEQKEQST 988
```

## RESULT 5

US-09-541-782-8  
Sequence 8, Application us/09541782  
Patent No. 6284480

## GENERAL INFORMATION:

APPLICANT: Nislow, Corey  
APPLICANT: Sakowicz, Roman  
APPLICANT: Beraud, Christophe  
TITLE OF INVENTION: Antifungal Assay  
FILE REFERENCE: 1015  
CURRENT APPLICATION NUMBER: US/09/541,782  
CURRENT FILING DATE: 2000-04-03  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8  
LENGTH: 1066  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-541-782-8

Query Match 5.1%; Score 119.5; DB 4; Length 1066;  
Best Local Similarity 19.1%; Pred. No. 0.03;  
Matches 96; Conservative 78; Mismatches 161; Indels 167; Gaps 21;

```
Qy 5 KQIEQLSKRPLSDSIC-----GVYLLKLSAERPLNENFNAQTALRKLSQNPASDER 58
Db 373 KEYTEEDIK-LKRDLMARDKNGIYLA-----EETVGEITVLKLESQREINERK 420
Qy 59 ----DAQQACLNKKILSD---SLYQO-----FSKTRDIE 88
Db 421 LILKALKDELQNEKIEFSEVSMVLKTEQELKKTENLNTKGTLLITKVKVLTTRRYK 480
Qy 89 LISMFVAQFLDPTLESANSLLEMLADLSEKHMDHLPVLP-----VETLKSDDKRE 143
Db 481 EKKEELVASHMKTPQVLTQAGQELAAADLATDTHQHLHGTIERRELDERKIRSCQGFND 540
Qy 144 REQADAKVKAFFOLVDSESSILVAPVLQPLVGEVTFPFDPSAERKGIISOLKSMLT 203
Db 541 RMQDN-----LEMIG-----GSLNLYQDQQAALKQLDSQ--EMVNS 574
```





Db 762 EATAAAMSAEODR-ENTRATLEQOULRSEERAAELASQLESTTAAKMSAEODRESTRAT 820

QY 249 FAKSLILRVENA-----LVHSGIKLAPKAEK----TVEQVAAESSVSEGELPSHMDT 298

Db 821 LEQOULRSEERAAELASQLESTTAAKMSAEODRESTRATLEQOULRSEERAAELASQLES 880

QY 299 KHERIPMASEQAQ-----TVSQHL 318

Db 881 TTAAK--MSAEODRESTRATLEQOUL 903

RESULT 8

US-08-282-845-2

; Sequence 2, Application US/08282845

; Patent No. 5719263

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; TITLE OF INVENTION: A 230kd Antigen Present in Leishmania

; TITLE OF INVENTION: Species

; NUMBER OF SEQUENCES: 3

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Immunex Corporation

; STREET: 51 University Street

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: Apple Macintosh

; SOFTWARE: Microsoft Word for Macintosh 5.1a

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/282,845

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/006,676

; FILING DATE: JANUARY 15, 1993

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Perkins, Patricia Anne

; REGISTRATION NUMBER: 34,693

; REFERENCE/DOCKET NUMBER: 5004-A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206)587-0430

; TELEFAX: (206)233-0644

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 955 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-282-845-2

Query Match 4.9%; Score 114.5; DB 1; Length 955;

Best Local Similarity 20.8%; Pred. No 0.074;

Matches 80; Conservative 56; Mismatches 150; Indels 99; Gaps 13;

QY 17 DDISICGVYLLKLSKSAFRLPNE-----FNVAQTALRKLSONPSADERDALQEACLKWK 70

Db 535 DAELASERKLESTVVAQLERQRRERVALDALQTHQKRLQEALESSESTAER----- 587

QY 71 ILSDSLYEQFS-----KTRDIELISM----- 92

Db 588 ---DQLLOQLTELOSERQTOLSQVVTDRERLRLDQRIQYEGETELARDVALCAQEMEA 644

QY 93 -FVAQFLLDTTLESANSLLEMLADLSEKHMDHLPVLPVETLKSDDDKGERQADAKV 151

Db 645 RYHAAYFHLQTLLELAT---EMEDALRERALAERDEAAALDAASTSONARESACERL 701

QY 152 KAFQVLGDSESSSILYAPVQLPLVGEVT-----FFDQSAERKGEISQLKSM 201

Db 702 TSLEQULRSEERAAELASQLESTTAAKSSAEODRENTATLEQOULRESEERAAELASQ 761

QY 202 TTVAQERFALQFKMEKAKCVNQ-----LDRLSALVSRKHSIGS-----QSTNFG 248

Db 762 EATAAAMSAEODR-ENTRATLEQOULRSEERAAELASQLESTTAAKMSAEODRESTRAT 820

QY 249 FAKSLILRVENA-----LVHSGIKLAPKAEK----TVEQVAAESSVSEGELPSHMDT 298

Db 821 LEQOULRSEERAAELASQLESTTAAKMSAEODRESTRATLEQOULRSEERAAELASQLES 880

QY 299 KHERIPMASEQAQ-----TVSQHL 318

Db 881 TTAAK--MSAEODRESTRATLEQOUL 903

RESULT 9

US-08-428-414A-3

; Sequence 3, Application US/08428414A

; Patent No. 5912166

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; TITLE OF INVENTION: LEISHMANIASIS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PC-DOS/MS-DOS

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/428,414A

; FILING DATE: 21-Apr-1995

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Kadlecsek, Ann T.

; REGISTRATION NUMBER: 39,244

; REFERENCE/DOCKET NUMBER: 210121.407

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; TELEX: 3723836 SEEDANDBERRY

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 955 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; US-08-428-414A-3

Query Match 4.9%; Score 114.5; DB 2; Length 955;

Best Local Similarity 20.8%; Pred. No 0.074;

Matches 80; Conservative 56; Mismatches 150; Indels 99; Gaps 13;

QY 17 DDISICGVYLLKLSKSAFRLPNE-----FNVAQTALRKLSONPSADERDALQEACLKWK 70

Db 535 DAELASERKLESTVVAQLERQRRERVALDALQTHQKRLQEALESSESTAER----- 587

QY 71 ILSDSLYEQFS-----KTRDIELISM----- 92

Db 588 ---DQLLOQLTELOSERQTOLSQVVTDRERLRLDQRIQYEGETELARDVALCAQEMEA 644

QY 93 -FVAQFLLDTTLESANSLLEMLADLSEKHMDHLPVLPVETLKSDDDKGERQADAKV 151

Db 645 RYHAAYFHLQTLLELAT---EMEDALRERALAERDEAAALDAASTSONARESACERL 701

QY 152 KAFQVLGDSESSSILYAPVQLPLVGEVT-----FFDQSAERKGEISQLKSM 201

Db 702 TSLBOQLRESEBAAELASOLEATAAKSSAEODRENTATLEOQLRESEBAAELASOL 761  
QY 202 TTYVAOEFATOFKMNKRCVTO-----LDRLSALVSTCHSGS-----OSTNFG 248  
Db 762 EATTAAKMSAEODR-ENTRATLEOQLRDESEBAAELASOLSTTRAKKSAEODRESTRAT 820  
QY 249 FAKSLTFRVENA-----LVHLSGIKLAPKAERK---TVEOEVAESSVSEGLPSHMDT 298  
Db 821 LEQOLROSEBAAELASOLESTTAAKMSAEODRESTRATLEOQLRESEBAAELASOLES 880  
QY 299 KHIERIPASQAOQ-----TVSQHL 318  
Db 881 TTAAK--MSAEODRESTRATLEOQL 903

## RESULT 10

PCT-US94-00324-1  
; Sequence 1, Application PC/TUS9400324  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven  
; TITLE OF INVENTION: Diagnosis of Leishmaniasis  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunex Corporation  
; STREET: 51 University Street  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Apple system 7.1  
; SOFTWARE: Microsoft Word, version 5.1a  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US94/00324  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/006,676  
; FILING DATE: 15-JAN-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Perkins, Patricia Anne  
; REGISTRATION NUMBER: 34,693  
; REFERENCE/DOCKET NUMBER: 5004-WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 587-0430  
; TELEFAX: (206) 233-0644  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 955 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US94-00324-1

Query Match 4.9%; Score 114.5; DB 5; Length 955;  
Best Local Similarity 20.8%; Pred. No. 0.074;  
Matches 80; Conservative 56; Mismatches 150; Indels 99; Gaps 13;

QY 17 DSDICGVYKLEKSAFRLRNE-----FNVAQTALRKLSQNPSPADERDALQOACLNKWK 70  
Db 535 DAEIASERKLESTVAQERQREREVALDALQTHQRKLQOALSSERTAAER-----587  
QY 71 ILSLSLYQFS-----KTRDIELISW-----92  
Db 588 ---DQLLOQLTLOSETLOSVYTDRELRTRDQRIQYEGETELARDVALCAQOEMEA 644  
QY 93 -FVAAGFLDITLSAANSLEWLADLSEKHMHLNPVLPVETLKSDDDKGKEREADAKV 151  
Db 645 RYHAAVFHLQTLLELAT---EMEDALRERALARDEAAAAAEIDAASSTSONARRESACRL 701

QY 152 KAFVOLGDSSESSITLAPVQLPLVGEVT-----FFDQSAERKGEISOLKSM 201  
Db 702 TSLBOQLRESEBAAELASOLEATAAKSSAEODRENTATLEOQLRESEBAAELASOL 761  
QY 202 TTYVAOEFATOFKMNKRCVTO-----LDRLSALVSTCHSGS-----OSTNFG 248  
Db 762 EATTAAKMSAEODR-ENTRATLEOQLRDESEBAAELASOLESTTAAKMSAEODRESTRAT 820  
QY 249 FAKSLTFRVENA-----LVHLSGIKLAPKAERK---TVEOEVAESSVSEGLPSHMDT 298  
Db 821 LEQOLROSEBAAELASOLESTTAAKMSAEODRESTRATLEOQLRESEBAAELASOLES 880  
QY 299 KHIERIPASQAOQ-----TVSQHL 318  
Db 881 TTAAK--MSAEODRESTRATLEOQL 903

## RESULT 11

US-09-310-187A-1  
; Sequence 1, Application US/09310187A  
; Patent No. 6358751  
; GENERAL INFORMATION:  
; APPLICANT: Benichou, Gilles  
; APPLICANT: Fedoseyeva, Eugenia  
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac  
; TITLE OF INVENTION: Graft Rejection  
; FILE REFERENCE: UCSF-090  
; CURRENT APPLICATION NUMBER: US/09/310,187A  
; CURRENT FILING DATE: 1999-05-12  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1939  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-310-187A-1

Query Match 4.9%; Score 114; DB 4; Length 1939;  
Best Local Similarity 21.9%; Pred. No. 0.26;  
Matches 93; Conservative 67; Mismatches 137; Indels 128; Gaps 20;

QY 7 QIEOLSKPL-----SDSICGVYK--LEKSAFRLRNE-----FNVAQTALR 47  
Db 1387 ELEEKKKKLAQRLQDAEAVAVNAKCSLEKTKHR-LQNEIEDLMVDERSNMAAALD 1445  
QY 48 KLSQNPSPADERDALQOACLNKWK-----ILSDSYEGFSKTRDIELISWFAAOFL 99  
Db 1446 KKQRN-----FDKTLAEMKQYEEOSELESSQKEANSSTELFKLNAYEESLKH 1496  
QY 100 LDTTLESANSLLEWLADLSE-----KHMHLNPV--LPVETLK-----SDDD 139  
Db 1497 LETFKRENKKNLQEFISDLTEQLGEGKVVHLEVKRQLEVEKLELQOALBEARASLEHE 1556  
QY 140 KGR-----BREQADAKVKAFFOLVGSSESSI-----LYAPVQL 174  
Db 1557 EGGLEIRAQLEFNOIKAEIERKLAKEDEMEQAKRNHQRVDSLOTSDAETRSRNEYLRV 1616  
QY 175 --PLVGEVTFPDFO-----SAPKGEISOLKSMLTTFVAO-----ERFAT 212  
Db 1617 KKKMEGDLNEMEIQLSIANRRAAEROKVKSLOSLAKDTQLOLDVAAYANDLKENAIV 1676  
QY 213 -----QFKMNKRCVTOQLDRLSAL-----VSTRCHSLGSQSTNFGAKSLTFRVE 258  
Db 1677 ERRNNLLQAELEELRAVVEQGERSKKLAEQELITSEKVVQLHQNLSLNQK---KME 1733  
QY 259 NALVHL-----SGIKLAPKAERK---TVEOEVAESSVSEGLPSHMDT-KHIERIPASQAO 312  
Db 1734 SDLTLOSEVEBAVOECHNADEK-AKKAITDAAMAELKKEQDTSAHLERMKKMEQTI 1792  
QY 313 TVSQHL 317  
Db 1793 KDLQH 1797



Db 346 ENT 348

RESULT 14  
US-08-751-965-9

; Sequence 9, Application US/08751965  
; Patent No. 5858360

; GENERAL INFORMATION:

; APPLICANT: Fukuda, Michiko N.

; TITLE OF INVENTION: Trophinin and Trophinin-Assisting

; TITLE OF INVENTION: Proteins

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; City: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/751,965

; FILING DATE: Herewith

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/439,818

; FILING DATE: 12-MAY-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-LA 2252

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 675 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-751-965-9

Query Match 4.7%; Score 110.5; DB 2; Length 675;

Best Local Similarity 21.2%; Pred. No. 0.1; Mismatches 130; Indels 103; Gaps 18;

Matches 77; Conservative 53; Mismatches 130; Indels 103; Gaps 18;

Db 291 HDGIVIKIEVQNDGSESLTPE--PLMG---OVEEHGFQDSBLXPCGEQPDLMQEP 345  
QY 454 QSS 456  
Db 346 ENT 348

RESULT 15

US-08-738-975-9

; Sequence 9, Application US/08738975

; Patent No. 5880267

; GENERAL INFORMATION:

; APPLICANT: Fukuda, Michiko N.

; TITLE OF INVENTION: Trophinin and Trophinin-Assisting

; TITLE OF INVENTION: Proteins

; NUMBER OF SEQUENCES: 22

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Campbell and Flores

; STREET: 4370 La Jolla Village Drive, Suite 700

; City: San Diego

; STATE: California

; COUNTRY: USA

; ZIP: 92122

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/738,975

; FILING DATE: herewith

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/439,818

; FILING DATE: 05-Dec-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Campbell, Cathryn A.

; REGISTRATION NUMBER: 31,815

; REFERENCE/DOCKET NUMBER: P-LA 2251

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 535-9001

; TELEFAX: (619) 535-8949

; INFORMATION FOR SEQ ID NO: 9:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 675 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-738-975-9

Query Match 4.7%; Score 110.5; DB 2; Length 675;

Best Local Similarity 21.2%; Pred. No. 0.1; Mismatches 130; Indels 103; Gaps 18;

Matches 77; Conservative 53; Mismatches 130; Indels 103; Gaps 18;

Matches 77; Conservative 53; Mismatches 130; Indels 103; Gaps 18;

QY 127 PVLPVETLKSDDDKGRKREQADAKVAFQVLVDSESSILYAVLQPLVGEVTFPFDQ 186

Db 56 PLPLKLRLKRPDPSPVPEETDMD-----PLOSPTSQKDTPOIS 96

QY 187 SAERK-----GEISQKSMITTVAOERFAIQ---FKME-----NAKR---CVTOLDR 228

Db 97 SGVQKEQPLPGETITRLGVMAVAQVERKLEAQMRLLTLEGRTGTEKKIADCEKTAVE 156

QY 229 LSAIVSTRCHSLGQSQINFGAKSLITRVENALVHLSG--IKLAPKAQVTEQVEAESS 286

Db 157 FANHLESKRWYVLGTLDEYGLQQRRLNMENTLKNRNFILRLPPG----- 202

QY 287 VSEGEIASHMDTKHIERIPMASEQAQTVSQHLHAGNISLGNLNNMRDLAFHLREVSD 346

Db 203 -SNEEVP-----KVPYTFDD---VAVHF---SEQEGNLSEWQKELYKNYMR--GN 244

QY 347 YFROSEPHSPISFLLEKAIKMGY--LSLPELLREM-----MSQNDALSTIFNAGLN 398

Db 245 Y-----ESLVSMDYAISKPDLSQMERGERPTMQQEDSEEGETPTDPSAA 290

QY 399 HLDQVLLPEVSTPTVIGIESPOTPAKPSVSDPKRVEEHVSQTS---PVDTQSKOD--QKP 453

Db 245 Y-----ESLVSMDYAISKPDLSQMERGERPTMQQEDSEEGETPTDPSAA 290

Mon Mar 3 10:44:16 2003

us-09-915-706a-2.raii

Page 9

Oy 399 HLDVLLPEVTPTVGIESPOTPAKRSYSDPRSVHEHSOTS----PVUTOSKOP-OKR 453  
| | | | | | | | | | | | | | | | | |  
Db 291 HDGIYIKREYGTNDGESSELETPE--PLMG---QVEEHGTFDSELGAPCGEQFDLDMQEP 345  
| | | | | | | | | | | | | | | | | |  
Oy 454 QSS 456  
| | | | | | | | | | | | | | | | | |  
Db 346 ENT 348

Search completed: March 2, 2003, 05:05:13  
Job time : 32 secs



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 04:19:19 ; Search time 23 Seconds

(without alignments)  
759.357 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343  
Sequence: 1 MFLSKHQIEQLSKPLSDSI.....DTSKQDQKQSSATSALSW 463

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174566 seqs, 37721826 residues

Total number of hits satisfying chosen parameters: 174566

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications -AA: \*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep: \*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep: \*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep: \*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep: \*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep: \*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep: \*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep: \*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep: \*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep: \*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep: \*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep: \*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep: \*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep: \*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	134.5	5.7	1179	10 US-09-815-242-13262	Sequence 13262, A
2	126	5.4	506	9 US-10-157-223-7	Sequence 11, Appl
3	126	5.4	610	9 US-09-952-267-11	Sequence 7, Appl
4	119.5	5.1	1179	10 US-09-815-242-13608	Sequence 13608, A
5	112	4.8	900	9 US-10-060-230-17	Sequence 17, Appl
6	110.5	4.7	900	9 US-10-060-230-18	Sequence 18, Appl
7	110.5	4.7	381	10 US-09-216-393-8	Sequence 8, Appl
8	110.5	4.7	476	10 US-09-779-307-16	Sequence 16, Appl
9	110.5	4.7	477	10 US-09-779-307-4	Sequence 4, Appl
10	110.5	4.7	1252	10 US-09-841-132-336	Sequence 336, App
11	109.5	4.7	824	9 US-09-884-001-2	Sequence 15, Appl
12	109	4.7	900	9 US-10-060-230-15	Sequence 15, Appl
13	109	4.7	900	9 US-10-060-230-16	Sequence 16, Appl
14	108.5	4.6	2125	10 US-09-919-172-29	Sequence 29, Appl
15	107.5	4.6	2025	10 US-09-815-242-5703	Sequence 5703, Ap
16	107.5	4.6	3158	10 US-09-815-242-12611	Sequence 12611, A
17	106	4.5	830	10 US-09-064-199-19	Sequence 19, Appl
18	106	4.5	1106	10 US-09-064-199-17	Sequence 17, Appl
19	106	4.5	1130	9 US-10-104-595-6	Sequence 6, Appl

20	106	4.5	1130	10 US-09-064-199-18	Sequence 18, Appl
21	106	4.5	1207	10 US-09-064-199-15	Sequence 16, Appl
22	106	4.5	1481	10 US-09-371-900-40	Sequence 40, Appl
23	106	4.5	1481	10 US-09-924-417-60	Sequence 60, Appl
24	105.5	4.5	2437	10 US-09-815-242-5834	Sequence 5834, Ap
25	105.5	4.5	6281	10 US-09-815-242-12996	Sequence 12996, A
26	105	4.5	1138	10 US-09-767-215-5	Sequence 5, Appl
27	105	4.5	2285	10 US-09-932-183A-2	Sequence 2, Appl
28	104.5	4.5	1369	10 US-09-729-674-42	Sequence 42, Appl
29	104.5	4.5	1548	9 US-10-025-380-1095	Sequence 1095, Ap
30	104.5	4.5	1348	10 US-09-922-217-1095	Sequence 1095, A
31	104	4.4	751	10 US-09-864-761-38419	Sequence 38419, A
32	104	4.4	1210	9 US-10-025-380-692	Sequence 692, App
33	104	4.4	1210	10 US-09-922-217-692	Sequence 692, App
34	104	4.4	1210	10 US-09-833-265-692	Sequence 692, App
35	103.5	4.4	331	9 US-09-987-107-38	Sequence 38, Appl
36	103.5	4.4	2478	10 US-09-815-242-5816	Sequence 5816, Ap
37	103.5	4.4	2478	10 US-09-815-242-12967	Sequence 12967, A
38	103	4.4	1884	10 US-09-785-770A-17	Sequence 17, Appl
39	103	4.4	1907	10 US-09-785-770A-16	Sequence 16, Appl
40	103	4.4	2472	10 US-09-815-242-5064	Sequence 5064, Ap
41	102.5	4.4	396	10 US-09-800-729-207	Sequence 207, App
42	102.5	4.4	829	10 US-09-946-805-8	Sequence 8, Appl
43	102.5	4.4	1019	10 US-09-801-574-76	Sequence 76, Appl
44	102.5	4.4	1499	10 US-09-911-826A-2	Sequence 2, Appl
45	102.5	4.4	1945	9 US-09-927-597-2	Sequence 2, Appl

#### ALIGNMENTS

RESULT 1  
US-09-815-242-13262  
Sequence 13262, Application US/09815242  
Patent No. US2002061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13262  
LENGTH: 1179  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-815-242-13262  
Query Match 5.7%; Score 134.5; DB 10; Length 1179;  
Best local Similarity 21.6%; Pred. No. 0.024;





```

RESULT 4
US-09-815-242-13608
: Sequence 13608, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zysek, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.01A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 13608
: LENGTH: 1179
: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
US-09-815-242-13608

```

RESULT 5  
 US-10-060-230-17  
 : Sequence 17, Application US/10060230  
 : Patent No. US20020173014A1  
 : GENERAL INFORMATION:  
 : APPLICANT: HILTONEN, Kaleervo  
 : TITLE OF INVENTION: A method to control cellular (3R)-hydroxyacyl-CoA  
 : TITLE OF INVENTION: esters, precursor molecules for polyhydroxyalkanoate  
 : TITLE OF INVENTION: synthesis in genetically modified organisms  
 : FILE REFERENCE: 0365-0528P  
 : CURRENT APPLICATION NUMBER: US/10/060,230  
 : CURRENT FILING DATE: 2002-02-01  
 : PRIOR APPLICATION NUMBER: 19991667  
 : PRIOR FILING DATE: 1999-08-03  
 : NUMBER OF SEQ ID NOS: 24  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ ID NO 17  
 : LENGTH: 900  
 : TYPE: FRT  
 : ORGANISM: *Saccharomyces cerevisiae*  
 : US-10-060-230-17

```

Query Match 4.8% Score 112: DB 9; Length 900;
Best Local Similarity 21.7%: Pred. No. 1;
Matches 84; Conservative 59; Mismatches 154; Indels 90; Gaps 16;

QY 2 PLSKHQIQE-----LSKPLSDSDSICGYLYLLEKSAFRLRNEFNVAQATALKLSQNPSPA 55
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 195 PLARSRMENVLPFILQLQGEKIVPLVLYTTHSTKSNISIFELAGFFQGLMWRSS 254
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 56 -----DERDALQELCKLKKWKILSDSLVEQSKTTRDIELISMFVAAGFLDITTESAN 109
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 255 GQIFNPDRKITYPPELILNKKKEITDYRDKPFKKTHPIQLSDY-----NDLITKAK 305
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
QY 110 SLEWLDLSEKHWHDLNPEVLETLKSDDDKKER-----EQADAKV-----KAFQLYGD 160
|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```



```

:
:
: PRIOR FILING DATE: 2000-02-10
: PRIOR APPLICATION NUMBER: 60/182,795
:
: PRIOR FILING DATE: 2000-02-15
:
: NUMBER OF SEQ ID NOS: 43
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 16
:
: LENGTH: 476
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
US-09-779-307-16

```

Query Match	4.7%;	Score 110.5;	DB 10;	Length 476;
Best Local Similarity	21.8%;	Pred. No. 0.52;		
Matches	78;	Conservative	53;	Mismatches 134;
			Indels	93;
			Gaps	15;

[illegible]

```

RESULT 9
US-09-779-307-4
; Sequence 4, Application US/09779307
; Patent No. US20020137675A1
; GENERAL INFORMATION:
; APPLICANT: Taupier Jr., Raymond
; APPLICANT: Majumder, Kumud
; APPLICANT: Verneil, Corine
; APPLICANT: Prayaga, Subhidas
; TITLE OF INVENTION: Polynucleotides and Polypeptides Encoded Therapy
; FILE REFERENCE: 15966-662 US
; CURRENT APPLICATION NUMBER: US/09/779,307
; CURRENT FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/180,880
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/181,044
; PRIOR FILING DATE: 2000-02-08
; PRIOR APPLICATION NUMBER: 60/181,656
; PRIOR FILING DATE: 2000-02-10
; PRIOR APPLICATION NUMBER: 60/182,795
; PRIOR FILING DATE: 2000-02-15
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 477
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-779-307-4

```

Query Match	4.7%;	Score 110.5;	DB 10;	Length 477;
Best Local Similarity	21.8%;	Pred. No. 0.52;		
Matches	78;	Conservative	53;	Mismatches 134;
			Indels	93;
			Gaps	15;

```

QY 27 LKSAFRPLRREFFVNAOTALRKLKSONSADERALDZCALNKKILSD---SLYEORSKT 83
Db 68 VNOSILSPMLNEVDPNIOAMRTOEE-----QIKTLNKFAFPYDKVRSLEOQNMKL 117
QY 84 TRDIELSMFEVAOFLDPTLTESAANLSEW-----LADLSEKH 121
Db 118 ETKSKLSLOOQOTARSNMNMNFESVNNLRKOLETGLGOEKLLEAEVLGMMOPLVEDERKNKY 177
QY 122 WDHNLPLVLPVET-----LKSDDDKKEREQOADAQAKAFQOLVGBSEESSILYAVLQDPLV 177
Db 178 EDELINKRTKEMENEVLILKKOAD-----EASNMKVELGSRLEGLDELNFL-----MQL--- 225
QY 178 GEVTFDFQSAERKEGELSOLKSMILTTVAOERAPLOKEMENAKRCYTQDLRLSALVSTKC 237
Db 226 -----YEEETIRELOQSIDPT-----SVLSDMDSN---CSLMDOSIIIAEVAAQY 265
QY 238 HSLGSGSTNGFCFASLSLITRVENALVHLSGIKRLAPKAAKTYVEDEVAESSYBEGELPSHMD 297
Db 266 EELANCSP--AEASMSMOIQIYEELQTLAGKHGDDRRTKTY--EISLTYGNISQLAHVD 320
QY 298 TKHTERIPMASEGOQVOTSOHLHA-----GNLSEL--GNULMNNRRLAPHLREYSD 346
Db 321 QR-----ASLEALITDAELHGLGVYKDAANAKLSELEALDLOAKQDMAQD--LREYOE 370

```

```

RESULT 10
US-09-841-132-396
: Sequence 396, Application US/09841132
: Patent No. US20020061848A1
: GENERAL INFORMATION:
:
: APPLICANT: Bhalla, Ajay
: APPLICANT: Sheikh, Jasir A.W.
: APPLICANT: Probst, Peter
: TITLE OF INVENTION: COMPOSITIONS AND METHODS
: FILE REFERENCE: 210121.469C8
: CURRENT APPLICATION NUMBER: US/09/841.132
: CURRENT FILING DATE: 2001-04-23
: NUMBER OF SEQ ID NOS: 599
: SOFTWARE: FastSeq for Windows Version 3.0/4.0.0
: SEQ ID NO 396
: LENGTH: 1252
:
: TYPE: PRN
: ORGANISM: Chlamydia pneumoniae
: US-09-841-132-396

```

	Query Match	4.7%	Pred. 110.5;	DB 10;	length 1252;
	Best Local Similarity	20.1%;	Pred. No. 2.1;		
	Matches	96;	Conservative	76;	Mismatches 169; Indels 137; Gaps 25.
Oy	23	VYLKLEKSAFPPLRNEFVNAOTALTKRLSQNSAD---	FRDAQOEACLNKKMK-----I	71	
	:	: : : :	: : : : : : : : : : : : : : : : : :		
Db	181	IYIHID----	KKKRRRKLTATFIIRALGYSSADDIIEEFFTGESSLSKEKFALLVGRI	236	
Oy	72	LSDSLYEDEFSTT--RDIELISMEVAOFLDPTTLLESANSLLEWLADSEKHMDHLNPVL	129		
	:	: : : : :	: : : : : : : : : : : : : : : : : :		
Db	237	LADNIDEASSLYYGKAGEKLS----	TAMLKRIMADIASVKIIVADDEHN-----	283	
Oy	130	PVEFTLSDDDKGKEREQADAKYKAFFOLVGDSSESIILA--PVQLPVGCEVTFFD---	184		
	:	: : : : :	: : : : : : : : : : : : : : : : : :		
Db	284	PIIKMLAND----	PTDSYEALAKDYYRRLRPOEPATYLLNARSTIMRL-----FFDPKR	332	
Oy	185	-----FOGAERKG-----EISOLSKMILTTPVAOERFAIOFMENAKRCVTOLDRLS	230		
	:	: : : : :	: : : : : : : : : : : : : : : : : :		
Db	333	YNIGRVGRTKLNRKLGFSIDDEALSQVILRKREDVIYALKYLIRLKMGEKACVDIIDHIA	392		
Oy	231	-----ALVSYTCHSLGSOSTINEGPAK-SLLTRVENALVHLSGIKLAPKAEAQTVBQE	281		
	:	: : : : :	: : : : : : : : : : : : : : : : : :		
Db	393	NRRVSVGELLIONOCRS-----GLARMKIYVERNNLTDFEDSDTLTP--GRVVSAAK	441		
Oy	282	-----VAESSVEGELPSHMDTKHIERITPMASEQAQTYSQHILAGNLSELGNLNNNRDLA	337		
	:	: : : : :	: : : : : : : : : : : : : : : : : :		
Db	442	GLASVYLKDIFEGRSQSLOPFMD-----QTNPAVELTHHKRLSLGP-GGINREBA	488		





```

Db 1291 DANKD---VDKQVQALIDEIGRNPLTDKEQALKDRINQILOQGH---NDINNALTKEA 1344
QY 194 ISOLKSMLTTFVAGERFAIOFKMENAKRCVQIDRLSALVSTKCHSLGSGOSTNFGFAKSL 253
Db 1345 IEQAKERLAQALQIDIKIDLKAK-EDAK-----NKIKALANAKRDQINSNP----- 1388
QY 254 LTRVENALVHLGKIKLAPKAEAKTVEQFVAESSVSEGLPSHMDPKHIERIPMASEQAQT 313
Db 1389 -----DLPEEQAKAKLKE-----IDEAE-----KRALQVENNAQT 1418
QY 314 VSQHLHAGNLSELGNLNNMNDLAFHLLREVSDFRQSEPHSPISFLEKAIKMGYLSLP 373
Db 1419 IDQ-LNRGLNLGLDDIRNTH-----VWEYDD-----QPAVNEISEATPEQLLVNG 1462
QY 374 ELL---REMSEONGDALSTIFNAGLNHLDOVLLPEVSTP-TVGIESPOTPOAKPSVSD 429
Db 1463 ELIYHRDDITEQ-DVL-----AHINLIDQLAEVITDPSTATISDSLAKVEVTLLD 1514
QY 430 PRSVEEHVSQTSPYDTOSKQ 449
Db 1515 GSKVIVNV---PVKVEKE 1530

```

Search completed: March 2, 2003, 05:44:03  
 Job time : 55 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 03:25:28 ; Search time 266 Seconds  
(without alignments)  
1122.223 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343  
Sequence: 1 MPLSKHQIQLEKSPLESDSI.....DTOSKODKQPSATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 64473110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_AA\_Main:\*

1: /cgn2\_6/ptodata/1/paa/US06.COMB.pep:\*

2: /cgn2\_6/ptodata/1/paa/US06.COMB.pep:\*

3: /cgn2\_6/ptodata/1/paa/US07.COMB.pep:\*

4: /cgn2\_6/ptodata/1/paa/US08.COMB.pep:\*

5: /cgn2\_6/ptodata/1/paa/US081.COMB.pep:\*

6: /cgn2\_6/ptodata/1/paa/US082.COMB.pep:\*

7: /cgn2\_6/ptodata/1/paa/US083.COMB.pep:\*

8: /cgn2\_6/ptodata/1/paa/US084.COMB.pep:\*

9: /cgn2\_6/ptodata/1/paa/US085.COMB.pep:\*

10: /cgn2\_6/ptodata/1/paa/US086.COMB.pep:\*

11: /cgn2\_6/ptodata/1/paa/US087.COMB.pep:\*

12: /cgn2\_6/ptodata/1/paa/US088.COMB.pep:\*

13: /cgn2\_6/ptodata/1/paa/US089.COMB.pep:\*

14: /cgn2\_6/ptodata/1/paa/US090.COMB.pep:\*

15: /cgn2\_6/ptodata/1/paa/US091.COMB.pep:\*

16: /cgn2\_6/ptodata/1/paa/US092.COMB.pep:\*

17: /cgn2\_6/ptodata/1/paa/US093.COMB.pep:\*

18: /cgn2\_6/ptodata/1/paa/US094.COMB.pep:\*

19: /cgn2\_6/ptodata/1/paa/US095.COMB.pep:\*

20: /cgn2\_6/ptodata/1/paa/US096.COMB.pep:\*

21: /cgn2\_6/ptodata/1/paa/US097.COMB.pep:\*

22: /cgn2\_6/ptodata/1/paa/US098.COMB.pep:\*

23: /cgn2\_6/ptodata/1/paa/US099.COMB.pep:\*

24: /cgn2\_6/ptodata/1/paa/US100.COMB.pep:\*

25: /cgn2\_6/ptodata/1/paa/US101.COMB.pep:\*

26: /cgn2\_6/ptodata/1/paa/US102.COMB.pep:\*

27: /cgn2\_6/ptodata/1/paa/US60.COMB.pep:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2343	100.0	463	23	US-09-915-706A-2
2	264.5	11.3	364	17	US-09-328-352-6641
3	197	8.4	681	16	US-09-252-991A-25690
4	152.5	6.5	1503	20	US-09-649-996-14
5	144	6.1	151	21	US-09-739-449-9942
6	144	6.1	151	22	US-09-803-110-9942

	7	138	5.9	2008	14	US-09-091-501-8	Sequence 8, Appl1
	8	138	5.9	2008	14	US-09-091-501B-8	Sequence 8, Appl1
	9	137.5	5.9	357	16	US-09-252-991A-17293	Sequence 17293, A
	10	136.5	5.8	5984	27	US-60-173-464-18560	Sequence 18560, A
	11	136.5	5.8	7118	27	US-60-167-217-22879	Sequence 22879, A
	12	136.5	5.8	7118	27	US-60-173-464-20674	Sequence 20674, A
	13	136.5	5.8	7182	27	US-60-191-637-39835	Sequence 39835, A
	14	136.5	5.8	7182	27	US-60-191-681-30862	Sequence 30862, A
	15	136.5	5.8	7201	20	US-09-614-150-40200	Sequence 40200, A
	16	135	5.8	372	22	US-09-897-516-7344	Sequence 7344, Ap
	17	135	5.8	372	27	US-60-215-161-7344	Sequence 7344, Ap
	18	135	5.8	2013	24	US-10-089-928-9	Sequence 9, Appl1
	19	134.5	5.7	1179	1	PCT-US02-03987-13262	Sequence 13262, A
	20	134.5	5.7	1179	21	US-09-752-069A-219	Sequence 219, Ap
	21	134.5	5.7	1179	22	US-09-815-242-13262	Sequence 13262, A
	22	134.5	5.7	1179	24	US-10-072-851-13262	Sequence 13262, A
	23	134.5	5.7	1179	27	US-60-174-089-219	Sequence 219, Ap
	24	133	5.7	1710	27	US-60-242-679-970	Sequence 970, Ap
	25	129.5	5.5	630	21	US-09-791-537-61506	Sequence 61506, A
	26	129.5	5.5	630	21	US-09-791-537-76922	Sequence 76922, A
	27	129.5	5.5	860	24	US-10-037-417-59	Sequence 59, Appl
	28	129.5	5.5	860	24	US-10-080-334-166	Sequence 166, Ap
	29	129.5	5.5	1036	27	US-60-161-932-977	Sequence 977, Ap
	30	129.5	5.5	1036	27	US-60-167-217-4618	Sequence 4618, Ap
	31	129.5	5.5	1036	27	US-60-171-627-366	Sequence 366, Ap
	32	129.5	5.5	1036	27	US-60-173-464-3755	Sequence 3755, Ap
	33	128.5	5.5	539	16	US-09-270-767-43655	Sequence 43655, A
	34	128	5.5	2017	1	PCT-US01-08631-36660	Sequence 36660, A
	35	127	5.4	1726	1	PCT-US02-21179-13	Sequence 13, Appl
	36	127	5.4	2278	1	PCT-US02-22866-14	Sequence 14, Appl
	37	126.5	5.4	752	21	US-09-791-537-13635	Sequence 13635, A
	38	126.5	5.4	752	23	US-09-917-234-70	Sequence 70, Appl
	39	126.5	5.4	752	25	US-10-171-311-48	Sequence 48, Appl
	40	126.5	5.4	824	25	US-10-108-260A-3569	Sequence 3569, Ap
	41	126.5	5.4	893	1	PCT-US02-31095-31	Sequence 31, Appl
	42	126	5.4	506	11	US-08-712-708-7	Sequence 7, Appl1
	43	126	5.4	506	25	US-10-157-223-7	Sequence 14, Appl1
	44	126	5.4	506	27	US-60-360-039-1648	Sequence 1648, Ap
	45	126	5.4	610	23	US-09-952-267-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1

US-09-915-706A-2

; Sequence 2, Application US/09915706A

; GENERAL INFORMATION:

; APPLICANT: NELSON, DAVID R.

; TITLE OF INVENTION: A LIVE, AVIRULENT STRAIN OF V. ANGUILLARUM THAT

; TITLE OF INVENTION: PROTECTS FISH AGAINST INFECTION BY VIRULENT V.

; TITLE OF INVENTION: ANGUILLARUM.

; FILE REFERENCE: 5112

; CURRENT APPLICATION NUMBER: US/09/915,706A

; CURRENT FILING DATE: 2001-07-26

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 463

; TYPE: PRT

; ORGANISM: Vibrio anguillarum

US-09-915-706A-2

Query Match	Score	DB 23:	Length	463:
Best Local Similarity	100.0%			
Matches	463:	Conservative	0:	Mismatches
			0:	Indels
			0:	Gaps

OY 1 MPLSKHQIQLEKSPLESDSI...GYYLLEKSAFPLNEFNVAOTLRKLSQNPSPADERDA 60

DB 1 MPLSKHQIQLEKSPLESDSI...GYYLLEKSAFPLNEFNVAOTLRKLSQNPSPADERDA 60

OY 61 LOEACLNKKKTLSDSLYEQFSKTRTDIELISFVAQAFLDITTLLESANSLMLADLSRK 120

Db 61 LOEACLNKWKILSDSYEOKSTTRTIELLSWFAAOFLLDTLESANSLLEWLADISEK 120  
QY 121 HMDHLNVLPEVETLKSDDDKEREQADAKVKAFFQVGDSEESSIIYAPVLOPLVGEV 180  
Db 121 HMDHLNVLPEVETLKSDDDKEREQADAKVKAFFQVGDSEESSIIYAPVLOPLVGEV 180  
QY 181 TFFDFOSEAKKEISQLSKMLTTTVAOERPAIOFKMENAKRCVTOIDRLSALVSTKCHSL 240  
Db 181 TFFDFOSEAKKEISQLSKMLTTTVAOERPAIOFKMENAKRCVTOIDRLSALVSTKCHSL 240  
QY 241 GSOSTNGCFKKSILTRRNALVHLSGIKLPAKPAKTEVEVASSVSEGLPSHMDTKH 300  
Db 241 GSOSTNGCFKKSILTRRNALVHLSGIKLPAKPAKTEVEVASSVSEGLPSHMDTKH 300  
QY 301 IERIPMAEQOQVSOHLHAGNLSLGNLNNMNDLAFHLLREVSDYFROSEPHSPISFL 360  
Db 301 IERIPMAEQOQVSOHLHAGNLSLGNLNNMNDLAFHLLREVSDYFROSEPHSPISFL 360  
QY 361 LEKAIKRGVLSLPELLREKMSSEONGDALSTFNAAGLNHLDOVLLPEVSTPTVIGIESPOT 420  
Db 361 LEKAIKRGVLSLPELLREKMSSEONGDALSTFNAAGLNHLDOVLLPEVSTPTVIGIESPOT 420  
QY 421 POAKPSVSDPRSVSEHVSQTSPTVTOQSKODOKPOSSATSALSW 463  
Db 421 POAKPSVSDPRSVSEHVSQTSPTVTOQSKODOKPOSSATSALSW 463

RESULT 2  
US-09-328-352-6641  
; Sequence 6641, Application US/09328352  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328, 352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 6641  
; LENGTH: 364  
; TYPE: PRF  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-6641

Query Match 11.3%; Score 264.5; DB 17; Length 364;  
Best Local Similarity 21.5%; Pred. No. 1,1e-14;  
Matches 87; Conservative 76; Mismatches 152; Indels 89; Gaps 14;

QY 8 IEQLSKPLSDSDSTICGVYKLEKSAFRPLRNEFNVAQATLRLKLSQNPASADERDALQEAOLN 67  
Db 5 ISELKIPINDSLCG-----EDYSF---SNEFHEIKKA--RTODDLDDQDWAERKQA 54  
QY 68 KWKILSDSYEOKSTTRTIELLSWFAAOFLLDTLESANSLLEWLADISEKHWHLNP 127  
Db 55 DWDFVAAVSSTLLIEKTKDRLTLWVLEAMTHLN-GFEGWVKGITLTHMLNQWODIHP 113  
QY 128 VLPETLKSDDDKEREQADAKVKAFFQVGDSEESSIIYAPVL--QLPVGSVTFE-- 183  
Db 114 IY-----EDDD-----LDQRIQLOGLINQ-----LPMLLKKVPLTNTAPYYNL 152  
QY 184 -----DFOSEAKKEISQLSKMLTTTVAOERPAIOFKMENAKRCVTO 225  
Db 153 LDYDNFLYHENIRRKQTEVEESQSGPSELDFQDAIFNT-----SKTFQSYNYQERNVS 206  
QY 226 LDRLSALVSTKCHSLGOSQTFNGFAKSLITRVENAL-----VHLSGIKLAKAKAKEYEQ 280  
Db 207 LTERNNVAKQITDHLMDGDSFPAIDSAFETIHSTLKKIYKAFAFGGLASQOAAV-- 264  
QY 281 EVASSVSEGLPSHMDTKHIERIPMAEQOQVSOHLHAGNLSLGNLNNMNDLAFHL 340  
Db 265 -----ITTPSMENQVYVQVSDQPMFQPAQOT---HL-----ANREQAMKV 302  
QY 341 LREVSDYFROSEPHSPISFLLEKAIKRGVLSLPELLREKMSSEON 384

Db 303 LOEIADYFOANEPHSPVSYMLQKTIKWSQMPHEMLAQVITKDEH 346

RESULT 3  
US-09-252-991A-25690  
; Sequence 25690, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252, 991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094, 788  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25690  
; LENGTH: 681  
; TYPE: PRF  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25690

Query Match 8.4%; Score 197; DB 16; Length 681;  
Best Local Similarity 20.9%; Pred. No. 5e-08;  
Matches 87; Conservative 79; Mismatches 156; Indels 94; Gaps 16;

QY 1 MPLSKH--QIEQLSKPLSDSDSTICGVYKLEKSAFRPLRNEFNVAQATLRLKLSQNPASADER 58  
Db 318 LPLSGNALSLLEVLLEPDPGQPCGPPSLRYDPDYDR-----LRELREDDSSLP 365  
QY 59 DALOEACLNK--WKILSDSYEOKSTTRTIELLSWFAAOFLLDTLESANSLLEWLAD 116  
Db 366 TGVQOAEAKRADAAYQIASLDLQRRSKDMLAMWGEA-WLQRGSLGLOALVYLAE 424  
QY 117 LSEKHWHLNVLPEVETLKSDDDKEREQADAKVKAFFQVGDSEESSIIYAPVLOPL 176  
Db 425 LCERYPEVHP-----QADGDGQSMRVPRIDMLLRKRYAELH-----TRLPL 466  
QY 177 VG-----EYTFEDFQSEAK-----GEISQLSKMLTTTVAOERPAIOFKMENAKR----- 221  
Db 467 MCGQFAFEITLYAQWRLQROQVAGDSKAKALEAAQLOOK-----KIDAEALRAEPLVQ 521  
QY 222 -----CVQIDRLSALVSTKCHSLGOSQTFNGFAKSLITRVENALVHLSGIKLP 271  
Db 522 WOKQASLACQOQLOQLREQWCD-RC--LQELAPSCQPLREVIQAQWLALKEF--IAMHP 576  
QY 272 KAEAKTEVEVASSVSEGLPSHMDTKHIERIPMAEQOQVSOHLHAGNLSLGNLNN 331  
Db 577 QAPLPEQPPVAEADAEG-----DADGESVVASAPSGPAGP----- 615  
QY 332 MNRDLAHLREVSDYFROSEPHSPISFLLEKAIKRGVLSLPELLREKMSSEONGDA 387  
Db 616 TSREDAYROLILLADYARTAREPHSPVPLIKRAVEGNKPLSELLLELI---NADS 668

RESULT 4  
US-09-649-996-14  
; Sequence 14, Application US/09649996  
; GENERAL INFORMATION:  
; APPLICANT: Plowman, Keith E.  
; TITLE OF INVENTION: KINASE GENES AND USES  
; NUMBER OF SEQUENCES: 53  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.



ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/649,996  
FILING DATE: 29-Aug-2000  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/976,255  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Marburg, Richard J.  
REGISTRATION NUMBER: 32,327  
REFERENCE/DOCKET NUMBER: 229/182  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-649-996-14

Query Match 6.5%: Score 152.5; DB 20; Length 1503;  
Best Local Similarity 21.4%: Pred. No. 0.0023; Mismatches 189; Indels 183; Gaps 25;

Matches 122; Conservative 75; Mismatches 189; Indels 183; Gaps 25;  
QY 3 LSKHQIEQLSKPLSD---DSICGVLYLLEKSAFRLPNEFNVAQTALRLKLSQNPASDERD 59  
DB LKPKOLEQ---PYSDRWYEVLOFCWLSPEK---RPAEDVHRLTYLTRLSQSDS---EVD 421  
QY 60 ALQEAFLNWKILISLSYQFSKTTTDLIELISFVAQF-----LLDTTLESANSLEM 113  
DB 422 FREQ-----WNLKPTNSRSDSNNAAPILDFARDRLGREMEVLTWTETSSQGLSFEY 476  
QY 114 LADLSE-KHMD-----HLN-----PVLPEFLKDDDKGREQADAVKAFPL 157  
DB 477 VWEAAHDFDEKSGHLDGLSYSTFYRVEYFESSLSDPGPKQ-----VOELSEN----- 522  
QY 158 VGDSSESSILYAPVQLPLVGEVTFPD-----FQSAERKQ---EISQLKSMLT 203  
DB 523 -DSDGDDVLRVP-----GVVPVDANHLNLSGSDYITQLEKSSNLELDYPPALLTT 574  
QY 204 TYAQERFAIQFWENAKRCVTQDLRLSALVSTKCHSLGSOSTNFGFAKSLTRVENALVH 263  
DB 575 -----QMDNBERGCPRLSQTALRSVE---LEESSTDEDFQSTDPKSSLPQ 620  
QY 264 LSGIKAPKAELKTYQEAESVSSGELPSHMDTKHIERIPMASQAQ-----TVSQHLH 319  
DB 621 DLHVTSGPSPPNNIPNDVQKSE---DLPSHQ--KIFDLMELNGVQAQFKPATLSSSLD 674  
QY 320 -----ACNLSELGNLNMNRRDLAFHLREVVDYFRQSE 352  
DB 675 NPKESVITGHFEKEKPKRIFFDSEPLCLSDNLHMQDNFPLN-----VOELSEN----- 723  
QY 353 PHSPIFLLEKAIKRWGYLSLPELLREMSQ-ONGDALSTIFNAAGLNHLN----- 401  
DB 724 -----FLQEKMLKGLSSLSKEHINDLQTEKNAQFTFAMLETSCNSLDTLQFAENKP 778  
QY 402 -----QVLL-----PEYSTPVIGISPPPTPAKPSVSPRSV 433  
DB 779 GUSLLOENVSTYGDPTDVMLTGDTLSTSLQSSSEVQVPPTSFTEETPRVPPSLPTQG 838  
QY 434 EEHVSTSPVDIOSKODKQSSATSALS 462

Db 839 E---TQPTCLDVIYVEDCLHQDISPDVAT 864

RESULT 5  
US-09-739-449-9942  
Sequence 9942, Application US/09739449  
GENERAL INFORMATION:  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15490)C  
CURRENT APPLICATION NUMBER: US/09/739,449  
CURRENT FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: US 09/514,000  
PRIOR FILING DATE: 2000-02-23  
NUMBER OF SEQ ID NOS: 13351  
SEQ ID NO 9942  
TYPE: prt  
ORGANISM: Agrobacterium tumefaciens  
US-09-739-449-9942

Query Match 6.1%: Score 144; DB 21; Length 151;  
Best Local Similarity 26.8%: Pred. No. 0.00027;  
Matches 44; Conservative 26; Mismatches 56; Indels 38; Gaps 5;

QY 222 CYTQDLRLSALVSTKCHSLGSOSTNFGFAKSLTRVENALVHLSGIKAPKAELKTY--- 278  
DB 4 CUSFPAITAVLSERCGQAAPPSSN-----IRNLTLEA-----AAIRTLGGR 46  
QY 279 EDEVAESSVSEGLPSHMDTKHIERIPMASQOQTVSQHLHAGNLSELGNLNMNRRDLAF 338  
DB 47 DDEPAP-----VEQTPAIR--AGTDESGOSAKRTSPASPEGISSHDEAF 88  
QY 339 HLREVSDFYRQSEPHSPISFLLEKAIKRWGYLSLPELLREMSQ 382  
DB 89 ETLISVARYFRTRERPHSPISLSTIELTVLRGRDMFSELLAELLPE 132

RESULT 6  
US-09-803-110-9942  
Sequence 9942, Application US/09803110  
GENERAL INFORMATION:  
APPLICANT: Hinkle, Gregory J.  
APPLICANT: Slater, Steven C.  
TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof  
FILE REFERENCE: 38-10(15490)D  
CURRENT APPLICATION NUMBER: US/09/803,110  
CURRENT FILING DATE: 2001-03-12  
PRIOR APPLICATION NUMBER: US 09/739,449  
PRIOR FILING DATE: 2000-12-19  
PRIOR APPLICATION NUMBER: US 09/514,000  
PRIOR FILING DATE: 2000-02-23  
PRIOR APPLICATION NUMBER: US 60/168,139  
PRIOR FILING DATE: 1999-12-01  
NUMBER OF SEQ ID NOS: 13351  
SEQ ID NO 9942  
LENGTH: 151  
TYPE: prt  
ORGANISM: Agrobacterium tumefaciens  
US-09-803-110-9942

Query Match 6.1%: Score 144; DB 22; Length 151;  
Best Local Similarity 26.8%: Pred. No. 0.00027;  
Matches 44; Conservative 26; Mismatches 56; Indels 38; Gaps 5;

QY 222 CYTQDLRLSALVSTKCHSLGSOSTNFGFAKSLTRVENALVHLSGIKAPKAELKTY--- 278  
DB 4 CUSFPAITAVLSERCGQAAPPSSN-----IRNLTLEA-----AAIRTLGGR 46  
QY 279 EDEVAESSVSEGLPSHMDTKHIERIPMASQOQTVSQHLHAGNLSELGNLNMNRRDLAF 338  
DB 47 DDEPAP-----VEQTPAIR--AGTDESGOSAKRTSPASPEGISSHDEAF 88

Db 47 DQBPAP-----VEOTPAIA--ACTDSGGSARTSPASPCIGSRDEAF 88  
QY 339 HLRVSDYFROSEPHSIFILEKAIKRWGLSLPELIREMSE 382  
Db 89 ETLISVAFRFRTERPHSISLIEFIVRGRMDSELLAEELLE 132

## RESULT 7

US-09-091-501-B  
; Sequence 8, Application US/09091501  
; GENERAL INFORMATION:  
; APPLICANT: Tinsley, Jonathon M  
; APPLICANT: Davies, Kay E  
; TITLE OF INVENTION: utrophin gene expression  
; FILE REFERENCE: 620-42  
; CURRENT APPLICATION NUMBER: US/09/091,501  
; EARLIER APPLICATION NUMBER: PCT/GB96/03156  
; EARLIER FILING DATE: 1996-12-19  
; EARLIER APPLICATION NUMBER: GB 9525962.8  
; EARLIER FILING DATE: 1995-12-19  
; EARLIER APPLICATION NUMBER: GB 9615797.9  
; EARLIER FILING DATE: 1996-07-26  
; EARLIER APPLICATION NUMBER: GB 9622174.2  
; EARLIER FILING DATE: 1996-10-24  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 2008  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric  
US-09-091-501-8

## Query Match

Best Local Similarity 5.9%; Score 138; DB 14; Length 2008;  
Pred. No. 0.08;  
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;

QY 1 MPLSKHQIEOLSKPLSDSICGVYIKLEKSAFRLNRFNVAQTALRKLSQNPSPADERDA 60  
Db 420 MELQKQLOQLSSWLA-----LTERIQKMSPPGLGDLPS 455  
QY 61 LQKACLNKWKILSDSLYEQSKTTRDIELT-----SMFVAQFLDPTTLESANSLFWLA 115  
Db 456 LQK-LLOEHKSLQNDLAEQVKVNSLTHMVYIVDENGESATLLEQOLKIGERWTAVC 514  
QY 116 DLSEKHHNDLNPVLPVETLKSDDDKGKERQDAKVKAFQVLVDSESSITVAPVQLP 175  
Db 515 RWTERRNRL-----OEISILMOELLEQ 538  
QY 176 LVGEVTFEFDQSAERKGEISQLKSMLTITVAQERFAIOFKMENAKRCVQIDRLSAL-- 232  
Db 539 CLLEAWLKEKEEALNKVQTSNFKQKELSVYRLATILKEDEMKR--QTLDOJLSEIGOD 596  
QY 233 -----VSTKCHSLGSGSTNGFPAKSLITRVENALVHLSGIKLAPKAAKTVQDEVA 283  
Db 597 VQQLSNPKASKKNNSSEELTORW--DSLVRLEDS-----SNQVTOAVA 640  
QY 284 ESSVSEGEELSHMDTKHIERIPMASBOA-----TVSOH-LHAGNLSSELGNLNNMRDL 336  
Db 641 KLGSQIPQKDLLETIVAVREKGMVKKPKQELPRPLTKAEHAKMRSTTELG----- 691  
QY 337 AFHLREVSDFROSEPH-----SPISFLEKAIKRWGLSLPELIREMSEONGDA 387  
Db 692 --ENLQELRLDQLOMEVHAKEIKWLNTLEMLSDKS-----LSLPE--RDKISE--S 738  
QY 388 LSTFNAAGLNHLDQVLLPEVSTPTVIGIESPQTPQAKPSVSDRPSVEHVSQTSPPVDQTS 447  
Db 739 LRTV-----NMTWKNKICREV--PTTLKECIOEPS--SVSQTR-IAAHNPVQKVVLLVSS 786  
QY 448 KODQPOSSATSALS 462

Db 787 ASDIPVOSHRTSEIS 801

## RESULT 8

US-09-091-501B-8  
; Sequence 8, Application US/09091501B  
; GENERAL INFORMATION:  
; APPLICANT: Tinsley, Jonathon M  
; APPLICANT: Davies, Kay E  
; TITLE OF INVENTION: utrophin gene expression  
; FILE REFERENCE: 620-42  
; CURRENT APPLICATION NUMBER: US/09/091,501B  
; EARLIER APPLICATION NUMBER: PCT/GB96/03156  
; EARLIER FILING DATE: 1996-12-19  
; EARLIER APPLICATION NUMBER: GB 9525962.8  
; EARLIER FILING DATE: 1995-12-19  
; EARLIER APPLICATION NUMBER: GB 9615797.9  
; EARLIER FILING DATE: 1996-07-26  
; EARLIER APPLICATION NUMBER: GB 9622174.2  
; EARLIER FILING DATE: 1996-10-24  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentln Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 2008  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; NAME/KEY: misc\_feature  
; LOCATION: (239) ... (230)  
; OTHER INFORMATION: Description of Artificial Sequence: Chimeric; Xaa = Unknown  
US-09-091-501B-8

## Query Match

Best Local Similarity 5.9%; Score 138; DB 14; Length 2008;  
Pred. No. 0.08;  
Matches 105; Conservative 72; Mismatches 172; Indels 146; Gaps 19;

QY 1 MPLSKHQIEOLSKPLSDSICGVYIKLEKSAFRLNRFNVAQTALRKLSQNPSPADERDA 60  
Db 420 MELQKQLOQLSSWLA-----LTERIQKMSPPGLGDLPS 455  
QY 61 LQKACLNKWKILSDSLYEQSKTTRDIELT-----SMFVAQFLDPTTLESANSLFWLA 115  
Db 456 LQK-LLOEHKSLQNDLAEQVKVNSLTHMVYIVDENGESATLLEQOLKIGERWTAVC 514  
QY 116 DLSEKHHNDLNPVLPVETLKSDDDKGKERQDAKVKAFQVLVDSESSITVAPVQLP 175  
Db 515 RWTERRNRL-----OEISILMOELLEQ 538  
QY 176 LVGEVTFEFDQSAERKGEISQLKSMLTITVAQERFAIOFKMENAKRCVQIDRLSAL-- 232  
Db 539 CLLEAWLKEKEEALNKVQTSNFKQKELSVYRLATILKEDEMKR--QTLDOJLSEIGOD 596  
QY 233 -----VSTKCHSLGSGSTNGFPAKSLITRVENALVHLSGIKLAPKAAKTVQDEVA 283  
Db 597 VQQLSNPKASKKNNSSEELTORW--DSLVRLEDS-----SNQVTOAVA 640  
QY 284 ESSVSEGEELSHMDTKHIERIPMASBOA-----TVSOH-LHAGNLSSELGNLNNMRDL 336  
Db 641 KLGSQIPQKDLLETIVAVREKGMVKKPKQELPRPLTKAEHAKMRSTTELG----- 691  
QY 337 AFHLREVSDFROSEPH-----SPISFLEKAIKRWGLSLPELIREMSEONGDA 387  
Db 692 --ENLQELRLDQLOMEVHAKEIKWLNTLEMLSDKS-----LSLPE--RDKISE--S 738  
QY 388 LSTFNAAGLNHLDQVLLPEVSTPTVIGIESPQTPQAKPSVSDRPSVEHVSQTSPPVDQTS 447  
Db 739 LRTV-----NMTWKNKICREV--PTTLKECIOEPS--SVSQTR-IAAHNPVQKVVLLVSS 786  
QY 448 KODQPOSSATSALS 462  
Db 787 ASDIPVOSHRTSEIS 801

```
RESULT 9
US-09-252-991A-17293
; Sequence 17293, Application US/09252991A
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO: 17293
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17293
```

```
Query Match          5.9%; Score 137.5; DB 16; Length 357;
Best Local Similarity 21.6%; Pred. No. 0.0046;
Matches 91; Conservative 67; Mismatches 145; Indels 119; Gaps 18;
```

```
QY 15 LSSDSICGYVTLKESAFRLNENFVAQTALRKLSO-NSPADERALDEACLNKKILS 73
DB 24 VSPDSCGDDL-----ETDAFLLELRIAGOGPEROMGAVLPAPPEMPRVR 71
QY 74 DSYEOPSKTRDIELISMFVAOFLDITLESANSLLEMLADISEKHNDLNPVLPVET 133
DB 72 ALASELFGH-SKDLRVANLLQSNVALD-GDGLADGLLVYRLLQYNDGYVPL----- 124
QY 134 LMSDDKKEKREDAQAKVAFQVLGDSSESSITLAPVQL--PLVEGVTF--FDPOS 187
DB 125 LADDD-----NDPFRIALTGLVAE-----PLDLYWAIVLVSRAGPYNLRA 170
QY 188 AERKGEISOLKSMLTITTVAOERFAIOFKMENAKRCVTQDRLSALVSTCHSGSOSTNF 247
DB 171 ALNAGLQRFAS---ETLSPEQIAGAFADADA-----DALATRAALEGAQEHAL 217
QY 248 GFASKSLITREVENA-----LVHL-----SGIKLAPKAEKATVEQVAES 285
DB 218 AIESGVAERVSAQGLDGLRQLRLQALQVFDLYGPGAGERLACGAEAADEQGA-A 276
QY 286 SVSEGLPSHMDTKHIERIPMASEQAQTVSQHLHAGNLSELGNLNMNRDLAEFHLLREVS 345
DB 277 PYAAVAAPA-----PRAS-----GEIA-----NREDVLRQDRL 306
QY 346 DYFROSEPHSPISFILEKAIKRCYLSLPFLLRMSEQNDALSTFNNAAGLNHLDOVLL 405
DB 307 EYIVRHEPSSPVVLLKRAKTLVTADFAIVRNLIPD-----GISQFETLRG 353
QY 406 PE 407
DB 354 PE 355
```

```
RESULT 10
US-60-173-464-18560
; Sequence 18560, Application US/60173464
; GENERAL INFORMATION:
```

```
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED G-PROTEIN COUPLED RECEPTORS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING GPCR PROTEINS AND USES
; FILE REFERENCE: CLO00173
; CURRENT APPLICATION NUMBER: US/60/173,464
; CURRENT FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 30269
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18560
```

```
; LENGTH: 5884
; TYPE: PRT
; ORGANISM: Drosophila
US-60-173-464-18560
```

```
Query Match          5.8%; Score 136.5; DB 27; Length 5884;
Best Local Similarity 22.0%; Pred. No. 0.69;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```
QY 30 SARPLRLNE-FNVAQTA---LRKLS-----QNPASD---ERP---ALDEA---CLNKK 70
DB 1259 SEFVDKNDIFSMLQTAETELRSLTLPQDPPKVVSDLSKRLDNLVQLQASHQLPKLH 1318
QY 71 ILSDSYEOPSKTRDIELISMFVAOFLDITLESANSLLEMLADISEKHNDLNPVLPVET 124
DB 1319 ALKSELAPLAAPDKRPI-LEKEVTEVEKMFNTMEHVKDVGYLEDSAK-MNNYTRRLA 1376
QY 125 -----LNPLP--VETLKSDDDKKEREQADAKVAFQVLGCD-SESSITLAPVQL-P 175
DB 1377 ELQEMANKVAPKNIKIALQSEDLTPREVRV---VKVQAFKRILGDGRMKQDLDLADASELAP 1433
QY 176 LGEVTFEFDQSAERKGEISOLKSMILT----- 203
DB 1434 KEGNIA-----EAKRLKGEITKLQEVLSAINRVNDAQAVQEDLVNMQFOAGLQOIKPA 1489
QY 204 -----TVQERFAIOFKMENAKRCVTQDRLSALVSTCHSGISQ 243
DB 1490 VEOSEVKNVNVSKPISLEBAVAMQOAOFEYO-----COEQDILHG-ISNISHKMLCK 1544
QY 244 STNFGFASKSLITREVEN-ALVHLSGIKLAPKAE-----KTVEQVYA--ESSVSEGE--- 291
DB 1545 -TN---APELDAMHSMRTAVHENAQASAKLEKLVANKSPFADAAKLEDMWGQGEQOM 1600
QY 292 --LPSHMDTKHIERI-----PMASEQAQTV-----SQHLHAGNLSEL-- 326
DB 1601 SRRPAVLNTPPHIDKLEKEVLKLSFNNEISQOQAKLVTLQGNADQISLHAPGGAALKD 1660
QY 327 -----GNLNMNRDLAFLHILREVSDYF--ROSEPHSPISFILEKAIKRCYLSLP 378
DB 1661 RVNOMKGLQKLEAVTRGH-TNEVSDAIIISRODPNAKLVE-----SNW-----MQLRN 1709
QY 379 MMS---EQNGDALST---IFNAGLNHLDO-----VLLPEVSTPYGIESPOTPOAKPSV 427
DB 1710 QVTQVEINERETSLHVIHALQEHADKKPFSNMIYDEVKDALG-----ATPESNAL 1765
QY 428 SDP-----RSVEEHVSQTSVPY-----DTOS-----KODKPOSSATSAL 461
DB 1766 NDAYTALVYVQNLFTNMLQKKAALBKWTELLGKKNDTESHLYLKHQLDKPEGAAEEL 1825
QY 462 S 462
DB 1826 PE 1826
```

```
RESULT 11
US-60-167-217-22879
; Sequence 22879, Application US/60167217
; GENERAL INFORMATION:
```

```
; APPLICANT: LI, Peter W. D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING DROSOPHILA PROTEINS, AND USES
; FILE REFERENCE: CLO00152
; CURRENT APPLICATION NUMBER: US/60/167,217
; CURRENT FILING DATE: 1999-11-24
; NUMBER OF SEQ ID NOS: 23195
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22879
; LENGTH: 7118
; TYPE: PRT
; ORGANISM: Drosophila
US-60-167-217-22879
```



```
Db 1323 ALKSELAPLADPKRPI-LEKEVTEVEKMFENTMEHVKDRVGLYEDYSAK-WNNYKTRLA 1380
QY 125 -----LNPLP---VETLKSDDDKGKEREOADAKVKAFFOLVGD-SESSITYAPVLQ-L-P 175
Db 1381 ELQEMANKVAPKNIELOSEDILTPEERV---VKVOAFKRLIGDRMKQDILLADASLAP 1437
QY 176 LVGEVTFEFDOSAERKEISQLKSMILT----- 203
Db 1438 KEGNIA-----EAKRLKEITKLOEVLASINRNVHQAOAVOEDLVNMQOFOAGLQIKPA 1493
QY 204 -----TVAQERFATQFKMENAKRCVTOIDRLSALVSTKCHSLGSO 243
Db 1494 VQOSEVKNVNVSKPISLEEAIVAMQOAOFFETQ---COEOLDKLHG-ISMNISHKMLCK 1548
QY 244 STNFGAKSLTRIVEN--ALVHLSGIKLAPKAE-----KTYQOEVA--ESSVSEGE--- 291
Db 1549 -TN---APDELDAMHSRMTAVHENAQAASAKLEKLVANMKSFDADAKLEDWVGQGEQOM 1604
QY 292 --LPSHMDTKHIERI-----PMASEQOQTV-----SOHLHAGNLSLSEL-- 326
Db 1605 SRPVALNTPHIDKLEKELVKLKSFNNEISQOQAKLVITGQANADQISLHAPGGAALAD 1664
QY 327 -----GNLNMRDLAFHLREVSDYF--ROSEPHSPISFLEKAIIRNGYLSLPELRE 378
Db 1665 RVNQMGKLOKLESEATRGH-INEVSDAISRODFNAKLVNF-----SNM-----MEQLRN 1713
QY 379 MMS-----EONGDALST---IFNAAGLNHLDO-----VLPEVSTPTVIGISPTOPKAPSV 427
Db 1714 QYQVEIENPERVETSLAHVILHLOEHADKPSFNAIYDEVKOLALG---ATPEESNML 1769
QY 428 SDP-----RSVEEHVSQTSVPV-----DTOS-----KODOKPOSSATSAL 461
Db 1770 NDAVYALVNVYQNLFTNMLQKKALEKMTLLGKWKNDTESHLYLKHQLDKPGPAEEL 1829
QY 462 S 462
Db 1830 S 1830

RESULT 14
US-60-191-681-30862
; Sequence 30862, Application US/60191681
; GENERAL INFORMATION:
; APPLICANT: LI, Peter, W.D.
; TITLE OF INVENTION: ISOLATED DROSOPHILA PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING ISOLATED DROSOPHILA PROTEINS AND
; TITLE OF INVENTION: USES THEREOF.
; FILE REFERENCE: C1000390
; CURRENT APPLICATION NUMBER: US/60/191,681
; CURRENT FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 30973
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30862
; LENGTH: 7182
; TYPE: PRK
; ORGANISM: DROSOPHILA
US-60-191-681-30862

Query Match 5.8%; Score 136.5; DB 27; Length 7182;
Best Local Similarity 22.0%; Pred. No. 0.97;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```
QY 176 LVGEVTFEFDOSAERKEISQLKSMILT----- 203
Db 1438 KEGNIA-----EAKRLKEITKLOEVLASINRNVHQAOAVOEDLVNMQOFOAGLQIKPA 1493
QY 204 -----TVAQERFATQFKMENAKRCVTOIDRLSALVSTKCHSLGSO 243
Db 1494 VQOSEVKNVNVSKPISLEEAIVAMQOAOFFETQ---COEOLDKLHG-ISMNISHKMLCK 1548
QY 244 STNFGAKSLTRIVEN--ALVHLSGIKLAPKAE-----KTYQOEVA--ESSVSEGE--- 291
Db 1549 -TN---APDELDAMHSRMTAVHENAQAASAKLEKLVANMKSFDADAKLEDWVGQGEQOM 1604
QY 292 --LPSHMDTKHIERI-----PMASEQOQTV-----SOHLHAGNLSLSEL-- 326
Db 1605 SRPVALNTPHIDKLEKELVKLKSFNNEISQOQAKLVITGQANADQISLHAPGGAALAD 1664
QY 327 -----GNLNMRDLAFHLREVSDYF--ROSEPHSPISFLEKAIIRNGYLSLPELRE 378
Db 1665 RVNQMGKLOKLESEATRGH-INEVSDAISRODFNAKLVNF-----SNM-----MEQLRN 1713
QY 379 MMS-----EONGDALST---IFNAAGLNHLDO-----VLPEVSTPTVIGISPTOPKAPSV 427
Db 1714 QYQVEIENPERVETSLAHVILHLOEHADKPSFNAIYDEVKOLALG---ATPEESNML 1769
QY 428 SDP-----RSVEEHVSQTSVPV-----DTOS-----KODOKPOSSATSAL 461
Db 1770 NDAVYALVNVYQNLFTNMLQKKALEKMTLLGKWKNDTESHLYLKHQLDKPGPAEEL 1829
QY 462 S 462
Db 1830 S 1830

RESULT 15
US-09-614-150-40200
; Sequence 40200, Application US/09614150
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: C1000728
; CURRENT APPLICATION NUMBER: US/09/614,150
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40200
; LENGTH: 7201
; TYPE: PRK
; ORGANISM: DROSOPHILA
US-09-614-150-40200

Query Match 5.8%; Score 136.5; DB 20; Length 7201;
Best Local Similarity 22.0%; Pred. No. 0.98;
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;
```

```

QY 30 SAERPLANE-FNVAQTA---LRKLS-----QNP SAD---ERD---ALQEA---CLNKWK 70
Db 1263 SEFVDQKNDJFSMLQTAETELRSILTPLOTDPKNVSQDLKSKRDJLNVLOQASHQLPLRH 1322
QY 71 ILSDSLYEQFSKTRDIELISWFFVAQOFLDITLESANSLLEWJLADLSEKHWDH----- 124
Db 1323 ALKSELAPLAPDKRP1-LEKEVTEVEKMFENMEHYKDRVGLLEDYSK-WNNYKTRLA 1380
QY 125 -----LNPVLP--VETLKSDDDKGKEREQADAKYKAFQOLVGD--SEESSILYAPVLOL-P 175
Db 1381 ELQEMANKVAPKNIEALQSELTPEERV---VKYQAKFRITIGDRMKQDLADASELAP 1437
QY 176 LVGEVTFPFDQSAERKGEISOLKSMJTT----- 203
Db 1438 KEGNIA---EAKRLKGEITKLEVLISAINRNVDHQAQVQEDLVNMQQFOAGLQITKPA 1493
QY 204 -----TVQERFALQPKMENAKRCYTQDLRLSALVSTKCHSLGSO 243
Db 1494 VEQSEVKNVNVSPKPILEEAQVAMQANQOPEYQ---COBQDLKLHG-ISHKMLCK 1548
QY 244 STNFGFAKSLTTRVEN--ALVHLSGIKLAPKAEA-----KTVEQEVVA--ESSVSEGE--- 291
Db 1549 -TN---APDELDAMHSRWTAVHENAKQASAKLEKLVANWKSFPDADAKLEDMVQGEQOM 1604
QY 292 --LP SHMDTKHIERI-----PMASEQAQTV-----SOHLHAGNISEL-- 326
Db 1605 SRPRAVINTPHIDKLEKELVYKLSFNNEISQOQAKLVTLGONADQISLHAPGGAALKD 1664
QY 327 -----GNLNMNRDLEFHLIREVSDYF--ROSEPHSPISFLLEKAIKRWIGLSLPELLRE 378
Db 1665 RVNOMKGRLOKLEATRGH--INEVSDAIIISRODFNAKLVPF-----SNW-----MEQLRN 1713
QY 379 MMS---EONGDALST---IFNAAGLNHLDQ-----VLLPEVSTPTVGIESPOTPOAKPSV 427
Db 1714 QVTQVEELINPERVETSLHVIHALQEHADKKRPFNATYDEYKQALAG---ATPESNAL 1769
QY 428 SDP-----RSVEEHVSQTSVP-----DTQS-----KODOKPOSSATSAL 461
Db 1770 NDAYTALVNVNQNLETNMLQKKALEKWTELLGWNKNDTESHLYLTKHQLDKPEGPAABEL 1829
QY 462 S 462
Db 1830 S 1830

```

Search completed: March 2, 2003, 05:10:57  
 Job time : 283 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 04:11:48 ; Search time 42 Seconds

(without alignments)  
1027.381 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343

Sequence: 1 MRLSKHQIEQSKPLSDSI.....DNQSKQDQKPOSSATSLSW 463

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 494224 seqs, 93196565 residues

Total number of hits satisfying chosen parameters: 494224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Pending\_Patents\_AA\_New:\*  
1: /cgn2\_6/ptodata/2/paa/PCIT\_NEW\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/paa/US06\_NEW\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/paa/US07\_NEW\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/paa/US08\_NEW\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/paa/US09\_NEW\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/paa/US10\_NEW\_COMB.pep:\*  
7: /cgn2\_6/ptodata/2/paa/US60\_NEW\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143	6.1	884	US-10-179-766-10	Sequence 10, Appl
2	131	5.6	1329	PCIT-US03-01772-29	Sequence 29, Appl
3	129.5	5.5	860	US-10-072-012-838	Sequence 838, App
4	126.5	5.4	752	US-60-423-586-154	Sequence 154, App
5	126.5	5.4	752	US-60-427-194-154	Sequence 154, App
6	126.5	5.4	785	US-10-309-851-24	Sequence 24, Appl
7	126.5	5.4	796	US-09-724-676-58406	Sequence 58406, A
8	126.5	5.4	796	US-09-724-676-58406	Sequence 58406, A
9	126.5	5.4	810	US-09-724-676-58407	Sequence 58407, A
10	126.5	5.4	810	US-09-724-676-58407	Sequence 58407, A
11	126.5	5.4	810	US-09-724-676A-58408	Sequence 58408, A
12	126.5	5.4	810	US-09-724-676A-58408	Sequence 58408, A
13	126.5	5.4	1133	US-10-309-851-14	Sequence 14, Appl
14	126.5	5.4	1133	US-10-309-851-16	Sequence 16, Appl
15	126.5	5.4	1135	US-10-309-851-112	Sequence 12, Appl
16	121.5	5.2	2642	US-09-724-676-55851	Sequence 55851, A
17	121.5	5.2	2642	US-09-724-676A-55851	Sequence 55851, A
18	121.5	5.2	2722	US-09-724-676-55868	Sequence 55868, A
19	121.5	5.2	2722	US-09-724-676A-55868	Sequence 55868, A
20	121.5	5.2	2915	US-09-724-676-55849	Sequence 55849, A
21	121.5	5.2	2915	US-09-724-676-55850	Sequence 55850, A
22	121.5	5.2	2915	US-09-724-676A-55849	Sequence 55849, A
23	121.5	5.2	2915	US-09-724-676A-55850	Sequence 55850, A
24	121.5	5.2	2929	US-09-724-676-55846	Sequence 55846, A
25	121.5	5.2	2929	US-09-724-676A-55846	Sequence 55846, A
26	121.5	5.2	2940	US-09-724-676-55847	Sequence 55847, A

27	121.5	5.2	2940	US-09-724-676-55848	Sequence 55848, A
28	121.5	5.2	2940	US-09-724-676A-55847	Sequence 55847, A
29	121.5	5.2	2940	US-09-724-676A-55848	Sequence 55848, A
30	121.5	5.2	2995	US-09-724-676-55867	Sequence 55867, A
31	121.5	5.2	2995	US-09-724-676-55867	Sequence 55867, A
32	121.5	5.2	2995	US-09-724-676A-55866	Sequence 55866, A
33	121.5	5.2	2995	US-09-724-676A-55867	Sequence 55867, A
34	121.5	5.2	3009	US-09-724-676A-55863	Sequence 55863, A
35	121.5	5.2	3009	US-09-724-676A-55863	Sequence 55863, A
36	121.5	5.2	3020	US-09-724-676-55864	Sequence 55864, A
37	121.5	5.2	3020	US-09-724-676-55865	Sequence 55865, A
38	121.5	5.2	3020	US-09-724-676A-55864	Sequence 55864, A
39	121.5	5.2	3020	US-09-724-676A-55865	Sequence 55865, A
40	121.5	5.2	3259	US-10-229-059-7	Sequence 7, Appl
41	118.5	5.1	1896	US-09-950-084-6335	Sequence 6335, Ap
42	118	5.0	674	US-09-724-676-51054	Sequence 51054, A
43	118	5.0	674	US-09-724-676-51055	Sequence 51055, A
44	118	5.0	674	US-09-724-676A-51054	Sequence 51054, A
45	118	5.0	674	US-09-724-676A-51055	Sequence 51055, A

## ALIGNMENTS

```
RESULT 1
US-10-179-766-10
: Sequence 10, Application US/10179766
: GENERAL INFORMATION:
: APPLICANT: Kenyon, Cynthia
: APPLICANT: Apfeld, Javier
: APPLICANT: Dillan, Andrew
: APPLICANT: Garigan, Delia
: APPLICANT: Hsu, Ao-Lin A.
: APPLICANT: Lehrer-Graetzer, Josh
: APPLICANT: Murphy, Coleen
: APPLICANT: The Regents of the University of California
: TITLE OF INVENTION: Eukaryotic Genes Involved in Adult Lifespan Regulation
: FILE REFERENCE: 023070-119300S
: CURRENT APPLICATION NUMBER: US/10/179,766
: CURRENT FILING DATE: 2002-06-24
: PRIOR APPLICATION NUMBER: US 60/300,577
: PRIOR FILING DATE: 2001-06-22
: PRIOR APPLICATION NUMBER: US 60/301,052
: PRIOR FILING DATE: 2001-06-25
: PRIOR APPLICATION NUMBER: US 60/373,975
: PRIOR FILING DATE: 2002-04-18
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 10
: LENGTH: 884
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans
: FEATURE:
: OTHER INFORMATION: 11w-4 (long-lived worm protein), gene F45H10.4
US-10-179-766-10

Query Match      6.1% Score 143; DB 6; Length 884;
Best Local Similarity 20.7%; Pred. No. 0.00074;
Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25;

OY 27 LEKSARPRNPNFNVQTLRLKSNPSPADERDALOEACLNKV-----KILSDLYQFQSK 82
DB 218 LRDATTEARLIRRLALJLTATLSOR-----RERLKITSQIKROYVAONLFFQOOR 268
OY 83 TTRDIELISWYAAQFLDITLESANSLSEWLADLSEKIMHDLNLPVLPVETLKSDDDKK 142
DB 269 LTRPVEDLVTESSKLAVEAPEETTOEETTTDASE-----VTTKAVEATE 317
OY 143 E-REQADAVKAFQVLQGSSESSILYAVLQ-----LPLVGEVTFPD-PQ 186
DB 318 EYTEATEATEA---PVAATTKSSSEMHVNTIRIMIRASAKDLISKYVTLISEKFSLEFE 374
OY 187 SAERK-----GEISOLKSMILTITVAQ-----ERFAIOFKEMAKRCVQL 226
```

Db 375 LADKKTLLTSKDEKISSMAKLDLINEALSEKESGETEQMEKFEKPEKSELVAME 434  
QY 227 DRLSALVSTCHSLGSGSTNGFAKSLTR-----VENALVHLSGIKLAPKAFAKTVQOE 281  
Db 435 DKDTPRNVFTISDSIKHKAELAKLHTIVSRNVYAEANA-----IEKEVEYPRAEKKEVEED 491  
QY 282 ---VAESSVE--GELPSHMDTKHIERIPMASEQAOTVSQHLHAGLSLGLNINN--N 333  
Db 492 VKAAEKKKEKKRGP--MKIEKLEK--PVDK-----SENHELKVLDD 534  
QY 334 RDLAFLHLLREVSDYFROSEPHSPISFLEK---AIRMGYLSLPELLREMMSEQ---NGDA 387  
Db 535 KERLLVSEIKNTAEETKPRVE--SEKSEETVAID---DMPALEKESAEKKEETGEP 589  
QY 388 LSTIFNAGLNHLDOVLLPEVSTPVGIESPQTPQAKPSVSDPRSEVHEVSOFSVPDQO 446  
Db 590 TTT-----EAAVETTEASETPPEAKPELLS--NLBDVLTITTP--ETET 630  
QY 447 ---SKODQKQSSATS 460  
Db 631 IEGSGEREPTTSAPAA 647

RESULT 2  
PCT-US03-01772-29  
; Sequence 29, Application PC/TUS0301772  
; GENERAL INFORMATION:  
; APPLICANT: INCYTE GENOMICS, INC.; YUE, Henry;  
; APPLICANT: GRIFFIN, Jennifer A.; RICHARDSON, Thomas W.;  
; APPLICANT: TANG, Y. TOM; THANGAVELU, Kavitha;  
; APPLICANT: FORSYTHE, Ian J.; BECHA, Shanya D.;  
; APPLICANT: CHAWLA, Narinder K.; HAFALIA, April J.A.;  
; APPLICANT: SWARNAKAR, Anita; MARQUIS, Joseph P.;  
; APPLICANT: GORVAD, Ann E.; BAUGHN, Mariah R.;  
; APPLICANT: LU, Dyung Aina M.; ARVIZU, Chandra S.;  
; APPLICANT: KABLE, Amy E.; LEE, Soo Yeun;  
; APPLICANT: RAMKUMAR, Jayalaxmi; JIANG, Xin;  
; APPLICANT: JACKSON, Alan A.; KHARE, Reena;  
; APPLICANT: ELLIOTT, Vicki S.; BULLOCH, Sean;  
; APPLICANT: XU, Yuming; LEE, Sally;  
; APPLICANT: LEHR-MASON, Patricia M.  
; TITLE OF INVENTION: STRUCTURAL AND CYTOSKELETON-ASSOCIATED PROTEINS  
; FILE REFERENCE: PF-1326 PCT  
; CURRENT APPLICATION NUMBER: PCT/US03/01772  
; CURRENT FILING DATE: 2003-01-16  
; PRIOR APPLICATION NUMBER: US 60/350,702  
; PRIOR FILING DATE: 2002-01-18  
; PRIOR APPLICATION NUMBER: US 60/351,715  
; PRIOR FILING DATE: 2002-01-25  
; PRIOR APPLICATION NUMBER: US 60/357,402  
; PRIOR FILING DATE: 2002-02-15  
; PRIOR APPLICATION NUMBER: US 60/379,880  
; PRIOR FILING DATE: 2002-05-10  
; PRIOR APPLICATION NUMBER: US 60/381,599  
; PRIOR FILING DATE: 2002-05-17  
; PRIOR APPLICATION NUMBER: US 60/387,270  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: US 60/397,125  
; PRIOR FILING DATE: 2002-07-19  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: PERL Program  
; SEQ ID NO 29  
; LENGTH: 1329  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Incyte ID No: 7512382CD1  
PCT-us03-01772-29

Query Match 5.68; Score 131; DB 1; Length 1329;  
Best Local Similarity 22.0%; Pred. No. 0.016;

Matches 108; Conservative 69; Mismatches 182; Indels 132; Gaps 23;  
QY 3 LSKHOLKPLSDSICGVYTLKESAPRPRNEFNVAQTALRLKLSQNSPAD---RD 59  
Db 907 LKAHQVQACHNLKEASSAQFEEL-IVLKEKENELKRLKLEAMLKREESLQTLQD 965  
QY 60 ALQACLNNKKT--LSDSLYQFSKTTIDIELIS-----WFAAQPLDITTE 105  
Db 966 VQDENKLFKSOIDQLKQOANOQASSPPHLELLKVISEREKETISGLMNE-----LDLKD 1020  
QY 106 SAANSLEMLADLSKEMHDHNPVLPVETLQSD--DQKREDO-----ADAKYAF 155  
Db 1021 AVEHQKKNNDLREKKNMEALASTEKMLQDYKNTSKERQOQVEAVELEAKVLLKLF 1080  
QY 156 QLVQDSEESSILYAPVQLPLVGEVTFPQSAERKGEISQLKSMLTITVAORFALQFK 215  
Db 1081 PKV--SVPSNLSGEML-----HGEKKAK-----BQMGTSQSEEVKYLEHK 1121  
QY 216 MENAKRCVQLDLRLSALVSTKCHSLGSGSTNGFPAKSLTRVENALVHLSGIKLAPKAA 275  
Db 1122 LKEA-----DEMHTLQLECEKY-----KSLAETEGILQKLO----- 1154  
QY 276 KTYEQEVAESSYSGELPSHMDTKHIERIPMASEQAOTVSQHLHAGLSLGLNINNMRD 335  
Db 1155 RSVQEQEENKMKVAVDE--SHKTIKQMOSSPTSSBQ-----ELERLRSENKD 1198  
QY 336 LAFHLLREVSDYFROSEPHSPISFLEKA-----IRMGYLSLPELLREMMSEON 384  
Db 1199 I-----ENLRREH--LEMELEKAMERSTYVTEVRELKAQNLNLTTLRLTDQNE 1247  
QY 385 ----GDALSTIFNAGLNHLDOVLLPEVSTPVGIE--SPOTQAKPSVSDPRSEVHEVS 438  
Db 1248 ROKVAGD---LHQAQSLTLQSKIYKAGDTTVIENSQVSEBETSESEKETSVS--LN 1301  
QY 439 QTSVPDQSKO 449  
Db 1302 QTV---TQLOQ 1309

RESULT 3  
; Sequence 838, Application US/10072012  
; GENERAL INFORMATION:  
; APPLICANT: Tchernev, Velizar  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Zehusen, Bryan  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Li, Li  
; APPLICANT: Gangoli, Bsha  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Anderson, David W.  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Miller, Charles E.  
; APPLICANT: Gerlach, Valerie  
; APPLICANT: Taupier Jr, Raymond J.  
; APPLICANT: Gusev, Vladimirt Y.  
; APPLICANT: Coleman, Steven D.  
; APPLICANT: Wolenc, Adam R.  
; APPLICANT: Pena, Carol E. A  
; APPLICANT: Furtak, Katarzyna  
; APPLICANT: Grose, William M.  
; APPLICANT: Alsobrook II, John P.  
; APPLICANT: Lepley, Denise M.  
; APPLICANT: Riegler, Daniel K.  
; APPLICANT: Burgess, Catherine E.  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-258  
; CURRENT APPLICATION NUMBER: US/10/072,012  
; CURRENT FILING DATE: 2002-01-31  
; PRIOR APPLICATION NUMBER: 60/265,102  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/265,514



;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/265,517  
;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/265,412  
;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/265,395  
;; PRIOR FILING DATE: 2001-01-31  
;; PRIOR APPLICATION NUMBER: 60/266,406  
;; PRIOR FILING DATE: 2001-02-02  
;; PRIOR APPLICATION NUMBER: 60/266,767  
;; PRIOR FILING DATE: 2001-02-05  
;; PRIOR APPLICATION NUMBER: 60/267,057  
;; PRIOR FILING DATE: 2001-02-07  
;; PRIOR APPLICATION NUMBER: 60/266,975  
;; PRIOR FILING DATE: 2001-02-07  
;; PRIOR APPLICATION NUMBER: 60/267,459  
;; PRIOR FILING DATE: 2001-02-08  
;; Remaining Prior Application data removed - See file wrapper or PALM.  
;; NUMBER OF SEQ ID NOS: 1391  
;; SOFTWARE: Patentln Ver. 2.1  
;; SEQ ID NO: 838  
;; LENGTH: 860  
;; TYPE: PRT  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Description of Artificial Sequence: Myosin tail  
US-10-072-012-838

Query Match  
Best Local Similarity 5.5%; Score 129.5; DB 6; Length 860;  
Matches 98; Conservative 87; Mismatches 208; Indels 137; Gaps 17;

QY 7 QTEOISKPLSD-----DSICGYLTKLESAFRLPNEFVNAQTALR 47  
DB 139 QTEOISKPLSD-----DSICGYLTKLESAFRLPNEFVNAQTALR 197  
QY 48 KLSQNSADERDLOACLNKWKILSDSYEQFSKTRDIE-----LISFVAAQFLDPT 103  
DB 198 ELQRO-----LNDLTSQKSRLOSENSDLTRQLEEAQVSNLSKQSGLESO 244  
QY 104 LESANSLEWLADLSKHHDLNPLVPTLKSDDDKGREGQADAKVAFQVLGDSSE 163  
DB 245 LEAKKSTLE-----EESREKANTLQALQRLQLEHDLDSLREOLEESEAFAELE-----RQ 293  
QY 164 SSTLYAPVQLPLVGEVTEFQSAERKGEISQLKSMLTYYA----- 207  
DB 294 LSKANAELQOMR-----SKFESEGLRAEELBELKKNOKISELEEADEANAKCDSLE 348  
QY 208 -----ERFAIOFKMENAKRCVQTL-----DRLSALVSTKCHSIG 241  
DB 349 KTKSRLOSELEDLOLEERANMAASELEKKKFNFKILAEKRRKVDLQALDPTAGREAR 408  
QY 242 SOSITNGFAKSLITREYENLVHLSGIKLAPKAETVEQVAE--SSVSEGLPSHMDTK 299  
DB 409 NLSTE-----LFRKNELEELKQVLEALREKNMLQDEIHDTQLQEGGRNHELEK 461  
QY 300 HIERIPMASEQOATVSQHLAG-NLSELG-----NLNMNRLAHLLREVSDFYRQSE 352  
DB 462 ARRLAEAEDELOALEEAELEESVYLRQVELSOIRSEIERLLEKEEFENRTRK 521  
QY 353 PHSPTSLLEKAIKMGYLSLPELREPMSEONDALSTFNAAGLNHLDVLLPEVSTPT 412  
DB 522 NH-----QRAIE-----SLQATLEAETKGAFA-----SRKKKLEGGDINELE 559  
QY 413 VGIESPOTQAKPSVSDPRSVEEHVSQTSPTVDTQSKODKPOSSATSALS 462  
DB 560 IALDHAN---KANAQAQKNVKKYQOOVKELQTOVEEQARARADAEQOLA 605

RESULT 4  
US-60-423-586-154  
; Sequence 154, Application US/60423586  
; GENERAL INFORMATION:

;; APPLICANT: American Home Products Inc.  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING SCHIZOPHR  
;; FILE REFERENCE: AM101288L  
;; CURRENT APPLICATION NUMBER: US/60/423,586  
;; CURRENT FILING DATE: 2002-11-05  
;; NUMBER OF SEQ ID NOS: 234  
;; SOFTWARE: Patentln version 3.1  
;; SEQ ID NO: 154  
;; LENGTH: 752  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
US-60-423-586-154

Query Match  
Best Local Similarity 21.9%; Score 126.5; DB 7; Length 752;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KQIOELSKPLS-----DDSI-----CGVYLKLESAFRLPNEFVNAQTALR 48  
DB 160 KLEVEKLSKRLMALEKLEDAFNKSKQECYSKCN--LEKERMPTKQLOSELESKVRKE 217  
QY 49 LSONPSADERD--ALOACLNKWKILS-----DSLYEQFSKTRDIEELISFVAAQFL 99  
DB 218 LEAIESRLKTEFTLKEED-LTKLKLTVFVDEKRTMSKLTKEDEKLQAAASQLOVEON 276  
QY 100 LDT-----LESANSLEWLADLSKHHDLNPLVPTLKSDDDKGREGQADAKVAF 154  
DB 277 KYTTYTEKLIETKTRALKSKTDEEKM-----SVTKEDDKLKNKLA- 319  
QY 155 FQVLGDSSESIYAPVQLPLVGEVTEFQSAER---KGEISQLKSMLTYYAQAERA 211  
DB 320 -----EEKGNLDSRYNNL-----KNRLOSLEIEKDFLNKNLQDSGKSTALHGENNK 370  
QY 212 IQPKMENAKRCVQTLDRLSAL-----VSTKCHSIGOSTN-FGAKSLITREYENLVH 264  
DB 371 IKELSOEVRRLKLLKDKMAIEDDLKTEDEYETLERRRYANERDKAOFLSKELEHYKME 430  
QY 265 SGIKLAPKAA-----KTVEQVAVSSVSEGLPSHMDTKIERIP--MASEQOATVS 315  
DB 431 AKKLEKETSHQWLFKRLQEEAKS-----GHSREYDALK-EKIHETMATEDLICH 485  
QY 316 QHLHAGNLSELGNLNMNRLAFHLLREVSDFYRQSEPHSPISFLEKAIKMGYLSLPEL 375  
DB 486 QGDHSVCKKKKLNQOENRNDLG--RELENLKLERYRHRS---KSLR-----PSL 531  
QY 376 LREPMSEONDALSTFNAAGLNHLDVLLPEVSTPTVGIESPOTQAKPSVSDPRSV 435  
DB 532 -----NGRRISD-----POVFSKEVQTEAVNDEPPDYKSLIP-----LER 566  
QY 436 HVSQTSPTVDTQSKODKPOSSATSALS 463  
DB 567 AVINGOLYEBSENQDEDPNDEC-SVLSF 593

RESULT 5  
US-60-427-194-154  
; Sequence 154, Application US/60427194  
; GENERAL INFORMATION:  
; APPLICANT: American Home Products Inc.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING AND TREATING SCHIZOPHR  
; FILE REFERENCE: AM101228L  
; CURRENT APPLICATION NUMBER: US/60/427,194  
; CURRENT FILING DATE: 2002-11-19  
; NUMBER OF SEQ ID NOS: 234  
; SOFTWARE: Patentln version 3.1  
; SEQ ID NO: 154  
; LENGTH: 752  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-60-427-194-154

Query Match  
Best Local Similarity 5.4%; Score 126.5; DB 7; Length 752;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;



```

Db      530  QGHSYVLOKKLNOOENNRPDLG-----REINLTKELERYRHFS-----KSLR-----PSL 575
Qy      376  LREHMEQNDALSTIENAGLHMLDOLLPEVSTPTVYGIESPOTPGAKPSVDPRSVEE 435
      11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      576  -----NERRISD-----PQVFEKEVQTEAVDNPEPDYKSLP-----LER 610
Qy      436  HYSOTSPVDTQSKODOKFQSSATSALSM 463
      : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 : 11 :
Db      611  AIVINGQLYESENEDDEPDNEG-SVLSF 637

```

RESULT 8  
US-09-724-676A-58406

```

sequence 3640b, Application US/09124676a
GENERAL INFORMATION:
APPLICANT: Comugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Comugen
CURRENT APPLICATION NUMBER: US/09/724.676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: Patentln version 3.2
SEQ ID NO 58406

```

ORGANISM: Homo sapiens  
US-09-724-676A-58406

Query Match	5.48;	Score 126.5;	DB 5;	Length 796;
Best Local Similarity	21.9%;	Pred. NO. 0.018;		
Matches 111; Conservative	77;	Mismatches 197;	Indels 123;	Gaps 24

[illegible]

RESULT 9  
US-09-724-676-58407  
; Sequence 58407, Application US/09724676  
; GENERAL INFORMATION:

```

: APPLICANT: CompuGen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 CompuGen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 58407
: LENGTH: 810
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-724-676--58407

```

Query Match	5.4%;	Score 126.5;	DB 5;	Length 810;
Best Local Similarity	21.9%;	Pred. No. 0.019;		
Matches 111; Conservative	77;	Mismatches 197;	Indels 123;	Gaps 24;

```

OY      5 KHOJEOUSKPLS-----DDSI-----CGVYLKLEKGAFFLRNEFVAOTALRK 48
Dh      218 KLEVEKLSKRIAMLEKLEDFNFNSKOECSLACN--LEKERMTTQOLSOELESKLVRRIKE 275
OY      49 LSONP$ADERD--ALOEA$CNKKKIL$-----DSLYEOFSKTRDIELLSWEVYAAOFL 99
Dh      276 LEATESRLKTEFEPLKED-LTKLTKLTVMFEVDERTKMSERLKKTEBDKLOAASSOLOVEON 334
OY      100 LDDT-----LESAN$S$EMVLADSEKHMHDLNPLVPEVTLKSDDDKGKREOADA$KAF 154
Dh      335 KVTIVTEKLEIEIKRALK$T$DVEEKMY-----SYTKERDRLKAKKFA- 377
OY      155 FOLVGDSEESSILYAPVLOLPLVGEVTFPDFO$AER--KGEISOLK$MLTTTVAO$KFA 211
Dh      378 -----BEEKNDL$R$VNM$L-----KNRLOSLEAIEKDFLKNKLNDSG$STALH$QENYK 428
OY      212 IO$FME$NAK$QCVPLD$EL$AL$-----V$TKAISLOSOSTN-FGFAK$SLT$REVAHALY 264
Dh      429 IKEL$OEYERLKLKLD$MK$IEDDLKKTEDYEYTLTER$RANERD$KAQ$PL$KLELV$MKEL 488
OY      265 SGIRLAD$A$A-----KTV$OEVAE$SV$BEGELP$HMD$TKHIERIP--M$SEQAQ$TV$S 315
Dh      489 AKYKLA$K$ET$T$SHEOWL$FKLOE$EAK$S-----GHLS$REVD$AL-K$IH$EY$M$T$EDL$ICHL 543
OY      316 OHL$AGN$L$EL$G$N$NN$NR$LA$H$LR$E$V$D$Y$RO$SE$P$S$PL$LEK$A$IR$G$Y$LS$PEL 375
Dh      544 QGDH$SVL$OKK$LNDO$EN$R$NL$G---RETDNL$TKLEL$RI$H$S-----K$LR-----P$SL 589
OY      376 L$R$M$SE$ONG$DAL$ST$F$N$AGL$H$LD$O$V$LL$P$E$V$T$T$V$IE$S$P$OT$R$O$AK$P$V$S$D$P$RY$E 435
Dh      590 -----NG$RISD-----P$Q$V$S$KE$V$Q$TE$A$V$ND$E$P$DY$K$SLIP-----L$ER 624
OY      436 H$V$O$T$P$V$D$O$K$O$D$K$P$O$S$A$T$S$AL$S$W 463
Dh      625 AVINGOLYE$E$EN$D$ED$P$N$E$G--SVL$S$F 651

```

RESULT 10  
US-09-724-676-58408

Sequence 58408, Application US/09724676  
GENERAL INFORMATION:  
APPLICANT: CompuGen LTD  
TITLE OF INVENTION: Variants of alternative splicing  
FILE REFERENCE: 129181.4 CompuGen  
CURRENT APPLICATION NUMBER: US/09/724,676  
CURRENT FILING DATE: 2000-11-28  
NUMBER OF SEQ ID NOS: 97222  
SOFTWARE: PatentIn version 3.2

```

; SEQ ID NO 58408
; LENGTH: 810
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-724-676-58408

```

Query Match	5.48;	Score 126.5;	DB 5;	Length 810;
Best Local Similarity	21.98;	Pred. No. 0.019;		

Matches	111, Conservative	77, Mismatches	197, Indels	123, Gaps	24, ...
QY	5	KH0IQLSKPLS----	DDSI-----	CGYILKLEKSAFRLRNEFVNAQTALRK	48
Db	218	KLVEYKLSKRMALKEKLDAFNRKSKOECSYLCKN--	LEKERMTTKQLSOELESKLKRIE	275	
QY	49	LSQNSASBERD--ALQEQACLNKMKIKLS-----	DSLQEQFSKTRRDIELLSMFVAAQFL	99	
Db	276	LEAISRLEKTEFTFLKED-LTKLTKLTVMFVDERKTMSEKLKTKTEDKLQQAASSQLOVEON	334		
QY	100	LDYTT-----LESANSLSEMLADLSKSHMDHLNPVLVETLKSDDDDKGEKREGQADAKKVAFL	154		
Db	335	KVTYTKTEKLIEBTKRALKSKTDVEEKMY-----	SVTKRDDLKKNLKA-	377	
QY	155	FQVGDSESSILTAPVLIQPLVGEVTEFFDFOSAR--	KGEISOLKSMLTITVAQERPA	211	
Db	378	-----EEEKGNDLTSRVNML-----	KNRLQSLTAIEKDFLKNLNDQSGKSTALHQQENNK	428	
QY	212	IQFMENNAKRCVYQIDRLISAL-----	VSTKHSLSGSOSTN-FGPAKSLITRYENALVYL	264	
Db	429	IKELSQEYERLKLKLDKMAIEDLMTKTEDEKETLERRANRDRDAQFLSKLELHVKMEL	488		
QY	265	SGIKLAPKAAE-----	KTYQEOVAESSVSXGCELPSHMDTGHIRIP--	MASEQAQVTS	315
Db	489	AKKYKAEKTEFTHSEQWFLRKLQDEEAKS----	GHLISREVDALK-EKIHIEYMATEDLICHL	543	
QY	316	QHLAAGNLSLSELGNLNNMRDLAFHLRLRESDYFROSEPHSPISFLLEKATIRMGVYLSLPEL	375		
Db	544	QGDHSVLQKLLNQGNRRNDLG-----	RELENTLTKLEKRRHNS-----	PSL	589
QY	376	LREMASEONGDALSTIFNNAAGLNHLIDQVLLPEVSTPYGVIESPQPPQAKPVSVDPRSVIE	435		
Db	590	-----NGRRISD-----	POVFSKREVQTEAVDNBPDDKSLIP-----	LER	624
QY	436	HVSQTSPIVDQSKODQKQPQSSATSALSW	463		
Db	625	AVINGQLYESENQDEDDPDNEG-SVLSLF	651		

```

RESULT 11
US-09-724-676A-58407
: Sequence 58407, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724.676A
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 58407
: LENGTH: 810
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-724-676A-58407

```

Query Match	5.4%;	Score 126.5;	DB 5;	Length 810;
Best Local Similarity	21.9%;	Pred. No. 0.019;		
Matches 111;	Conservative 77;	Mismatches 197;	Indels 123;	Gaps 24;

[illegible]

Db	378	-----EEBKGDLLSRVMM-----KNRLOSLLEAKERFLANKLNQDSCKSTTALHQENNK	428
Qy	212	IOFMENNAKRCVTOIDRLISAL-----VSTKCHSLGSOSTN-FCGFAKSLTRYENALVHL	264
Db	429	IKELSOEYERIKLKLDMKAIEDDLMTKEDEYEFLERRYRNERDKAOFLSKELEHVMMEL	488
Qy	265	SGIKLAPKAA-----KTYCEVAARSSVEGELPSHMTKHTERIP--MASEAOATVS	315
Db	489	AKYLAETETSHCOWFLKRLQEEAKS-----GHLSSEVDALK-EKHTHEMATEDLCHL	543
Qy	316	QHLHAGNLSELGNLNNNRDLAFHLREVSDFRQSEPHSDISFLLEKAIWGLSLPEL	375
Db	544	QGDHSVLQKRLNQEENRRDGLG-----REIENLTKELERYRHS-----KSLR-----PSL	589
Qy	376	LREMMSEONGALSTIFNAAGLNHLDQYLLPEVSTPIYVGLSEPTQPAKRSVSDPRVEE	435
Db	590	-----NGRRISD-----POVFSKEYQTEAVDNEPDPYKSLIP-----LER	624
Qy	436	HVSQTSPYDTQSKODOKQSSATSALSM	463
Db	625	AVINQOLYESENODEDPNDG-SVLSF	651

```

RESULT 12
US-09-724-676A--58408
; Sequence 58408, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58408
; LENGTH: 810
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A--58408

```

Query Match	5.48;	Score 126.5;	DB 5;	Length 810;
Best Local Similarity	21.98;	Pred. No. 0.019;		
Matches 111; Conservative	77;	Mismatches 197;	Indels 123;	Gaps 24

Qy	5	KHOEJOLSKPIS-----DDSI-----	CGVYLKJESKSAFPLRNENFNAOZALRK	48
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :		
Db	218	KLEVEKUSKRIMALEKLEDDANNSKOCESYLKCN--	LEKERMHTKOLSOLESUKRIKE	275
Qy	49	LSONPSADERD--ALOEACLKMKILS-----	DSLYPEQSKTRIDIELISMFVAOFL	99
Db	276	LEAIESRLKEFTPLKED-LTKLKLTVPMFVDEKRT	TESEKILKEDTKQAASSOLOEON	334
Qy	100	LDIT-----LESAANSLEWLAJLSEKIMDHLPVL	PVPEYTLKSDDDKKEQADAKYKAF	154
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :		
Db	335	KVTYVTEKLIETETRALKSKTDVEBMY-----	SVYKREDDDKNKTKAKA-	377
Qy	155	FOLVGDSSESSILPAVYLOPLVEGVNFPROSAER---	KGEISOLKSMYLTITVAOERFA	211
Db	378	-----EEKGMDLISRYNML-----KNRLOSLEA	IEKREFPKNNKLNOBSGKSTTALHOCNNK	428
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :		
Qy	212	IQFMENAKRCVTOJDLRLSAL-----VSTRCHSL	SGOSTN-EGFAXSLTRVEMALVHL	264
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :		
Db	429	IKELSOEYERIKLTKOMKKALETEDBEYETLERRA	NYANERDKQFSLKELEHYKMWEL	488
Qy	265	SGIKLAPKAEK-----KTYVEQVAASSVSEGE	LPSHMDTKHIERIP--MASEQAOYVS	315
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :		
Db	489	AKYKLAETETSHQOWLFKRLQEEAKS-----GHL	SREVDALK-EKHEYMATEDLICHL	543
Qy	316	OHLNAGNISLSEGINNNNRDLAFLLREVSQY	PROSPHSPISPLEKATIMQYLSLPEL	375
Db	544	QGDHSVLQKTLNOCENNRDGG-----REIEUTK	LEERYHNF5---KSLR-----PSL	569

```

Oy 376 LREMMSENCALSTIFENAAGLNHLDOVLLPEVSTPRVIGIESPOTQAKSVSPREV 435
      || : || | | | | | | | | | | | | | | | | | | | | | | | |
Db 590 -----NCRISD-----PQVSENYOTEAUNDNPPYKSLIF-----LER 624
      || | | | | | | | | | | | | | | | | | | | | | | | |
Oy 436 HVSQTSPIVDTSRKODQKQFSATSALSWM 463
      : | | | | | | | | | | | | | | | | | | | | | | | |
Db 625 AVINGQLYESENODEDPNDEG-SVLSP 651
      : | | | | | | | | | | | | | | | | | | | | | | | |

```

## RESULT 13

```

US-10-309-851-14
: Sequence 14: Application US/10309851
: GENERAL INFORMATION:
: APPLICANT: Saus, Juan
: APPLICANT: Revert-Ros, Francisco
: TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
: TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
: FILE REFERENCE: 98.723-F-US
: CURRENT APPLICATION NUMBER: US/10/309,851
: CURRENT FILING DATE: 2002-12-04
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 14
: LENGTH: 1133
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-309-851-14

```

Query Match	5.48	Score	126.5	DB	6	Length	1133
Best Local Similarity	21.98	Pred. NO.	0.032				
Matches 111; Conservative	77	Mismatches	197	Indels	123	Gaps	24

Db	414	KLEVEKSTRIMALEKLEDAFNKSCOCESYLCKN--LEKERMTTKOLSOELESLEKVIKRE	4711
OY	49	LSQNSADERP--ALOEAACKUKKILS-----DSLYEOPSKTRBILISMFVAOFL	99
Db	472	LEAIESRLEKREFFLCKED-LTKLKTLYVMPEDEKRTSEKIKTEKEDKLOQAASSOLOVEON	530
OY	100	LDTT-----LESAA NSLEWLDLSEKHMDHLNPLYVETLKSDDDKGEREQADAKYKAF	154
Db	531	KVTYVTEKLIBEETRAKLSKTDVEEKMV-----SVYKREDDDKLNKKA-	573
OY	155	FOLVGDESESSILVAPYLOLPLVEGVFFPOSAER---KGEISOLKSMLTTPYVAOERA	2111
Db	574	-----EEBKGNDDLISRVNML-----KNPLQSLAELEKOPFLNKNLQDSCKSTTALHQNNK	624
OY	212	IOFMENAKRCVOTOLDRLSAL-----VSTKCHSLGSOSTN-EGFAKSLTLRVENALVHL	264
Db	625	IKELSOEVERTKLKLDIMKALIEDMLKTEDEYELFERRYANERNRKAOFLSKELEHYKMEI	684
OY	265	SGIKLAPKAAE-----KTYGEVAESSVSEGLPSHMOPKHIERIP--MASEQAOTVS	315
Db	685	AKYKLAETETESHQWLEFKRLQGEBAAS-----GILSSEVDAK-EKHETMATEDLICH	739
OY	316	QHLIAGNLSELGNLNMRDLAFHLLREVSDFROSEPHSPISFLLEKAIIRMGVLSPEL	375
Db	740	QGDHSVLOKKLINOQENRRDGG-----REIENTKLELERYRHS-----PSL	785
OY	376	LREMMSENGDALSTIFMAAGLNLDDOVLLEPVSTPIVGVIESPOTPAKSVSDPREVEE	435
Db	786	-----NGRRID-----POVSEKVEOTEAVDNPEPYKSLIP-----LER	820
OY	436	HVSOTSPPVDTOSKODOKPOSSATSLSM	463
Db	821	AVINGQLYEESENODEPDEPNDEG-SVLSFR	847

RESULT 14  
US-10-309-851-16  
; Sequence 16, Application US/10309851  
; GENERAL INFORMATION:

```

: APPLICANT: Saus, Juan
: APPLICANT: Revert-Ros, Francisco
: TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
: TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
: FILE REFERENCE: 98.723-F-US
: CURRENT APPLICATION NUMBER: US/10/309,851
: CURRENT FILING DATE: 2002-12-04
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 16
:
: LENGTH: 1133
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: US-10-309-851-16

```

Query Match	5.4%	Score 126.5	DB 6	Length 1133
Best Local Similarity	21.9%	Pred. No. 0.032		
Matches 111, Conservative	77	Mismatches 197	Indels 123	Gaps 24

[illegible]

```

RESULT 15
US-10-309-851-12
; Sequence 12, Application US/10309851
; GENERAL INFORMATION:
; APPLICANT: Saus, Juan
; APPLICANT: Revert-Ros, Francisco
; TITLE OF INVENTION: GIPs, a Family of Polypeptides with Transcription Factor Activity
; TITLE OF INVENTION: Interact with Goodpasture Antigen Binding Protein
; FILE REFERENCE: 98, 723-F-US
; CURRENT APPLICATION NUMBER: US/10/309, 851
; CURRENT FILING DATE: 2002-12-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1135
; TYPE: PRT
; ORGANISM: Homo sapiens

```

US-10-309-851-12

Query Match 5.4%; Score 126.5; DB 6; Length 1135;  
Best Local Similarity 21.9%; Pred. No. 0.032;  
Matches 111; Conservative 77; Mismatches 197; Indels 123; Gaps 24;

QY 5 KHQIEQLSKPLS-----DST-----CGVYLKLEKSAPRPLRNEFNVAQTALRK 48  
DB 414 KLEVERLSKRIMALEKLEDAFNKSKQECYSIKCN--LEKERMTYKQLSQELSLKVRKE 471  
QY 49 LSQNPASDERD--ALQACLNKKILS-----DSLVEQSKTTRDIELISMFVAQFL 99  
DB 472 LEAIESRLKEFTFLKED-LTKLTLTVMVEYDERKTMSEKLUKTEDKLQAASQLOVEON 530  
QY 100 LDFT-----LESAANSLEMLADLSEKHMHLNPLVPETLKSDDDKGKEREQADAKYAF 154  
DB 531 KVTVTVEKLIETTRALKSKTDVEKMY-----SVTKERDDLKKNLKA- 573  
QY 155 FQLVGDSEESSILYAPVQLPLVGEVTFPFQSAER--KGEISQLKSMLTITVAQERFA 211  
DB 574 -----EEKGNLDLSRYNML-----KNRLQSLAELEKDFLKNKLNQDSGKSTTALHOENNK 624  
QY 212 IQFKENAKRCVQOLDRLSAL-----VSTKCHSLGSOSTN-FGFAKSLTRVENALVHL 264  
DB 625 IKELISQEVERLKLKLDKMAKIEDDLMKTEDEYETLERRYANERDKAQFLSKLEHVKMEL 684  
QY 265 SGIKLAPKAE-----KTVEQVAAESSVSEGLPSHMDTKHIERIP--MASEQAQTVS 315  
DB 685 AKYKLAETKETSHEQWLFKRIQEEAKS---GHLSREVDAK-EKIHVMATEDLICH 739  
QY 316 QHLHAGNLSEIGLNINNNRDLAFHLRLREVSDFRQSEPHSPISFLLEKAIKRWGYSLEPEL 375  
DB 740 QGDHSVLQKKLNQOENNRDLG---REIENLTJKELEERYRHS---KSLR-----PSL 785  
QY 376 LREMMSONGDALSTIFNAGLNHLDQVLLPEVSTPYGIESPQTPQAKPSVSDPRSYEE 435  
DB 786 -----NGRRISD-----POVFSKEVOTEAVDNEPPDYKSLIP-----LER 820  
QY 436 HVSQTSFVDTQSKODKPOSSATSALSM 463  
DB 821 AVINGQLYESENQDEDDPNDG-SVLSF 847

Search completed: March 2, 2003, 05:41:33  
Job time : 59 secs

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 03:23:43 ; Search time 39 seconds  
(without alignments)  
1141.288 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343  
Sequence: 1 MFLSKHQIEQLSKPLSDSI.....DRGSKODKPKSSATSALSM 463

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241.5	10.3	351	AC3090	conserved hypothet
2	241.5	10.3	351	G98196	hypothetical prote
3	209	8.9	366	DB3350	hypothetical prote
4	152	6.5	2245	T18278	myosin heavy chain
5	143	6.1	884	T20405	hypothetical prote
6	141.5	6.0	344	G83634	hypothetical prote
7	141.5	6.0	1510	T16927	hypothetical prote
8	134.5	5.7	1179	E98012	hypothetical prote
9	134	5.7	1051	T18502	hypothetical prote
10	129.5	5.5	1133	T22976	apsb protein - Eme
11	127.5	5.4	2471	T42977	large tegument pro
12	126.5	5.4	1023	AE1643	ATP-dependent dsDN
13	126	5.4	506	S50914	cell division cont
14	125.5	5.4	1287	T22235	hypothetical prote
15	125.5	5.4	1992	A47297	myosin heavy chain
16	125	5.3	1194	T37503	probable chromosom
17	124.5	5.3	550	T40370	hypothetical prote
18	124.5	5.3	1940	S04090	myosin heavy chain
19	124.5	5.3	1940	A24922	myosin heavy chain
20	124.5	5.3	1964	A59282	nonmuscle myosin I
21	123.5	5.3	867	C71638	DNA polymerase I (
22	123.5	5.3	1051	T18351	Imp1 protein - Myc
23	123	5.2	1935	A37102	myosin beta heavy
24	123	5.2	1935	S06006	myosin beta heavy
25	123	5.2	2022	T43214	ovtl protein - nem
26	122.5	5.2	2469	H36812	hypothetical prote
27	122	5.2	1934	I48153	myosin heavy chain
28	122	5.2	2109	E89066	protein H05009.1
29	122	5.2	2109	T33247	hypothetical prote

30	121.5	5.2	1179	2	G95144	conserved hypothet
31	121.5	5.2	3225	2	I52300	giantin - human
32	121.5	5.2	3259	1	A56539	giantin - human
33	120.5	5.1	999	2	JC5278	oxygen-regulated p
34	120.5	5.1	1938	2	A59293	skeletal myosin he
35	120	5.1	650	2	S44806	F10B9.6 protein -
36	120	5.1	828	2	F96535	hypothetical prote
37	120	5.1	1034	2	T32297	hypothetical prote
38	119.5	5.1	1066	1	A48669	kinesin-related pr
39	119.5	5.1	1690	2	T13030	microtubule bindin
40	119	5.1	1023	2	AE1280	ATP-dependent dsDN
41	119	5.1	1356	2	S32763	kinectin 1 - human
42	119	5.1	1935	2	A59286	myosin heavy chain
43	119	5.1	2104	2	T38774	myosin-3 heavy cha
44	118.5	5.1	1920	2	A53188	pericentrin - mous
45	118.5	5.1	1937	2	I38035	myosin heavy chain

ALIGNMENTS

RESULT 1

AC3090 conserved hypothetical protein Atu4343 [imported] - Agrobacterium tumefaciens (strain

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #extl\_change 11-Jan-2002

C:Accession: AC3090

R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woc

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kulyavin, T.; Levy, R.; Li, M.; MCCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kan

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AC3090

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-351 <KUR>

A:Cross-references: GB:AE008689; PIDN:AA145137.1; PID:g17742810; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4343

A:Map position: linear chromosome

Query Match 10.3%; Score 241.5; DB 2; Length 351;

Best Local Similarity 23.6%; Pred. No. 2.3e-08;

Matches 96; Conservative 64; Mismatches 162; Indels 85; Gaps 15;

QY	5	KHOIEOLSKPLSDSDSICGVYTLKLE---KSAFRPLRNFENVAQOTALRKLSNPDAEDDAL	61
DB	7	KRDIEFLGN-----CGDNIRNDSKTRREIYRIKDRNRAR-----AEERAAAS	48
QY	62	QEAQL---NKKILSDSLVDFQSKTRTRDIELISGFVAAGFLDPTTLESAANSLEMLADLS	118
DB	49	PQDLKLSSSWDSVSNLGLDIYSESKDVELLMLAASLR-L-RGFHGLREIYELCGDLF	107
QY	119	EKHMDHLNPLVPVETTLKSDDDKGERQADAKKAPFOLVGDSESSILYAPVLQPLV-	177
DB	108	YNHMDSLRSI-----SDND-----BEKAPFAGLNGISSEGT-LVQPLRLASLIP	152
QY	178	---GEVTFDFQSAERKGEISOLKSLTTVAQERFAIOFKKEMNAKRCVTQIDRSALV	233
DB	153	GKGFGEISLMDFOLAQPNRSKRREELRYLASEGVAAAMSHLAAVTCLSSDPAITAVL	212
QY	234	STACHSLGSGSTNFGFAKSLLTRVENALVHLSGIKLAPKAERTV---EDEVASSVSEG	290
DB	213	SERCGAAPPSSN-----IRNTLIRBA-----AAITLGGRDGEPAP-----	249
QY	291	ELPSHMDTKHIERIPMASEAOQIVYSHLHAGNLSLGNLNNMRDLAFILLREVSDFRQ	350
DB	250	-----VEQTPALA---AGTDESGQSAARTSPASPESIGSSIDEAFETLLSVARVFR	297

```

QY      351 SEPHSPISFLLKAIWMGYLSPELLREMMSEONGDALSTIENAGL 3977
          :||||| : : : ||| :: | : : |||
Db      298 TEHPSPISLSIETLVRRGRMDFSELLAELLPEET--QARNAVLTAGI 3422

```

## RESULT 2

698196  
hypothetical protein AGR\_L1042 [imported] - *Agrobacterium tumefaciens* (strain C58, Ceratophyllaceae)  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: 698196  
R:Goodner, B.; Hinkley, G.; Gatungu, S.; Miller, N.; Blanchard, M.; Qurolojo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2338, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*  
#:Reference number: A97359; PMID:11743194

A:Accession:G98196  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-351 <R0>  
A:Cross-references: GB:AE007870; PIDN:AAK89097.1; PID:g15158901; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_1042  
A:Map position: linear chromosome

Query Match	10.38	Score 241.5	DB 2	Length 351
Best Local Similarity	23.68	Pred. No. 2.3e-08		
Matches 96	Conservative 64	Mismatches 162	Indels 85	Gaps 15

```
Oy      5 KHQIEOLSKPLSDSDSICGVYKLE--KSAFRPLRNEFNVAOTLRKLQSNPASADERDAL 61
          | || |           || :: : : : : || | | | : || | |
Db      7 KRDIETLGN-----CGDNIRNDSRTRETYIRIKDERNQAR-----AERAAS 48
```

```

Oy      62 QEACI--NKWKILSDSLYEQSKTTRDIELISWFAAQFLDITLESANSLEWLADLS 118
      : | : | : : : : : : : : : : : : : : : : : : : : : : : :
Db      49 PÖDNKISSSDSVSNGLÖIITISESKDVEILAMLAEASLRV-RGFHGLREIYELCGDLF 10

```

```

QY 119 EKHWDHLNPLVEVETLKSDDDKGKREQADAKVKAFFQLVGSSESSIIYAPVILQLPLV 17
      ||| | : ||| : | | | | : | : | :
Db 108 YNHWSLRSI-----SDDND-----EKFAPFAGLNGICSEGT-LVQPLRLASLIP 15

```

```

QY 178 ---GEVTFEPFQSAERKGEISOLKSMLT'TTVAQERRAIQFKMENAKRCVTLQDLRLSALV 23
      || : ||| || : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 GKGFEHSLMPQOLAQRPNESKREELRYTASAGVAAMSSHLAAVNTCLISFDLITAVL 21

```

```

QY 234 STCHSLGSGSTNFGFAKSLTLRYENALVHLSGIKLPAKAEKTV---EQEVAESSVSEG 29
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 213 SERCGOAPPPSN-----IRNTLLEAA-----AAIRTLGGRDQEPAP----- 24

```

```
OY      291 ELP SHMDTKHIERIPMASEAQTVSQHLHAGNLSSELGNLNNMRDLAFHLLRVSDFRQ   35
          :: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      250 -----VEOTPAIA--ACIDESGQSARTSPASPEGISSRDEAFETLLSVARYFR   29
```

```
QY      351 SEPSPISFLEKAIIRMGVLSLPPELLREMMSEQNGDALSTINAGL 397
        :||||||| : | : | : ||| :: | : : |||
Db      298 TEPSPDISLSTETLVRRGRMDESELLAEILPET--CARNAVLTAAGI 342
```

### RESULT 3

D83350

C:Accession: D83350  
C:Species: Pseudomonas aeruginosa  
C:date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an oppo  
A:Reference number: AB2950; MUID:20437337; PMID:10984043

A:Accession: D83350  
A:Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-366 <STO>  
A:Cross-references: GB:AE004662; GB:AE004091; NID:99948395; PIDN:AA605748.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2360

Query Match	8.9%	Score 209;	DB 2;	Length 366;
Best Local Similarity	21.5%	Pred. No. 3.1e-06;		
Matches 91; Conservative	73;	Mismatches 151;	Indels 108;	Gaps 16

```
Oy      1 MPLSKH--QIEOLSPKPLSDSICGVYLLKEKSAPRPLRNENFVNAQTALRLSONPSADER 58  
       :| | : | | : | : | : | : | : | : | : | : | : | : | : | : |  
Db      3 LPLSGNALSEVLLEPIPGGACGSPSLRYPPDYR-----LRELREDDSLP 50
```

QY 59 DALQACLNK--WKILSDSLYEQFSKTYTRDIELISWEVAAQFLLDTTESANSIEMLAD 11  
 : : : | :  
Db 51 TGVWAFAEAKRDWAAVEQLASSELLQRRSKDLMAANIWGEA-WLORGIGJGLQRALVLAE 10

```
QY 117 LSEKHMIDLNPVLPVETLKSDDDGKEREQADAKVAFAFQLVGDSSESSILYAPVLQPL 17
      |:::| |:::|:::|
Db 110 LCERPVEEVHP-----QAQGDGDSWRVPIDMLRRYAEHLH-----TRLPL 15
```

```
QY 177 VG-----EVTFFDQSAERK---GEISQLKSMLTTVAGERFAIQFMENAKR----- 22
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 152 MGOGAFAEITTLVAMORLOFOQVAGSDSKSAKALEAOLQOK----KIDALFAEPLVQ 20
```

```

QY 222 -----CYTQL-----DRLSALVSTKCHSLGSGSTNGFAKSLITRVENALVHL 26
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 207 WORKOASLACOOOLRLTEOWCDBCLGELAPSCOPREVE-----IAOWALLKETIAMH- 26

```

QY 265 SGIKLPKAEAKTVEQEVAESSVSEGLPSHMDTKHTERIPASEQAQTVSQHLAGNLS 32  
| : | : ||| : || | : : :  
Db 261 -----POAPLSEEPDPVAEADASEG-----DTGGESVPSPASPGAGAP----- 30

```
QY      325 ELGNLNMNRDIAFHLLREVSDYFRÖSEPHSPISFLLEKAIWGYLSPELLEMMSEQN 38
        :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      301 -----TSREAYRQLLTADYIARTPEPSPVPVYIKRAVEGNGKPISEIIAELL---N 35
```

QY	385	GDA	387
mb	351	ADS	353

## RESULT 4

### T19279

myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 08-Sep-2000  
C:Accession: M19279, M20570

R;Hammer III, J.A.; Jung, G.  
J. Biol. Chem. 271, 7120-7127, 1996  
A>Title: The sequence of the dictyostellium myo J heavy chain gene predict  
A;Accession number: 710854. WITH 0631510. PMID:063147

A:Accession: T18278  
A:status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A; Cross-references: EMBL:U02409; NID:g1150765; PTD:g1150766; PTDN:AAA0518  
R; Titus, M.A.; Kuspa, A.; Loomis, W.F.  
Proc. Natl. Acad. Sci. U.S.A. 91, 9446-9450, 1994

A:Reference number: Z20873; MUID:95023928; PMID:7937787  
A:Accession: T30579  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-190, 'F', 192-283, 'T', 285-290, 'R', 292-331, 'IE', 'WM', 337-338, 'A';  
A:Cross-references: EMBL:L35322; NID:q1039360; PID:q1039361; PIDN:AAA7985  
C:Genetics:

A:introns: 257/3, 307/3  
C:superfamily: myosin heavy chain, myosin motor domain homology  
C:keywords: ATP, P-loop





Db 264 PVAAYAPA-----PRAS-----GEIA-----NREDVLRQLRL 293

Qy 346 DYEROSPEHSPISFLLEKAIKRWGLSLPELLREMSSEONDALSTIFNAGLNHLDOVL 405

Db 294 EYVVRHEPSSPVVLLKRAKTLVTADPAEIVRNLIIP-----GISOFELRG 340

Qy 406 PE 407

Db 341 PE 342

RESULT 7

hypoetical protein T23F2.2 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T16927

R:Du, Z.

submitted to the EMBL Data Library, October 1995

A:Description: The sequence of C. elegans cosmid T23F2.

A:Reference number: Z18608

A:Accession: T16927

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1510 <DU2>

A:Cross-references: EMBL:U39649; NID:g1049370; PID:g1049372; PIDN:AAA80384.1; CESP:T23F2

C:Genetics:

A:Gene: CESP:T23F2.2

A:Introns: 20/2; 81/1; 119/2; 211/3; 236/3; 290/2; 344/1; 396/3; 439/3; 465/3; 556/3; 60

Query Match

Best Local Similarity 20.0%; Score 141.5; DB 2; Length 1510;

Matches 112; Conservative 89; Mismatches 168; Indels 191; Gaps 26;

Qy 9 EQLSKPLSDSICGVYIKLEKSAFRLRNFPNNAOTLRKLSQNSPADERDALOEALNK 68

Db 619 QOLSEHSD-----FLSTQOT-LHDLKKEYSAKNTTL-----VDKREVEILLAK 663

Qy 69 WKILSDSLVEQFSK-----TRDIELISMFVAAGFLDPTLESANSLFWLADLSEKHMDH 124

Db 664 TELV-DALVTQOLENIRKQDREL-----LKQSERDQYKKSLEMFIAEK----- 708

Qy 125 LNPVLPVETLKSDDKGKEREQADAKVK---AFQO-----LVGDS----- 161

Db 709 -VPLFEELIQLSKDK-----NEITARLKHQVEYFEDELAKLINDSMIKKREDDYLETNI 763

Qy 162 --EESSLVAPVQLPLVGEVTFPDSAPR-KGEISQLKSMLTITVAQERFALQFMEN 218

Db 764 RANESMT-----ERKLEISGLKKDLNOKMOAHLQ---KSEL 798

Qy 219 AKRCVTOIDRLSALVS-----TKCHSLGSGSTNGFAK----- 251

Db 799 EKRLLSDIHVSQOSRVNHSQRDVECOAIPROINKYVGCKPKNKNETIEKALPDEN 858

Qy 252 -----SLITRVENALVHL-----SGIKLAPKADEAKTVEQVAESS 286

Db 859 EERLRICKAELETTTRRVTVLQQLVSIIOQOSSQIKRKIAVVEDNSKNVTHTEDLESK 918

Qy 287 VSEGEPLSHMDTKHIERIPMASEQAOTVSOHLHAGNLSELGN-----LNNMR 334

Db 919 MKVEL--KNTELMERID-SLEAEFVAASIEKSRIOQLVNEPDNLKQKILNDMSVYSK 974

Qy 335 DLAF-----HLREVSDFROSEPHSPISFLLEKAIKRWGLSLPE-LIREMSEON-CD 386

Db 975 EKQWLQWRINLEKDNELQKOIOPSS-----EKSLK-----SLNKGLTRKTHSEPPFGD 1024

Qy 387 ALSTIFNAGLNHLDOVLPEVSTPTVGISSPQ-----TPQAKPSVSDPR 431

Db 1025 DMSTEGGASTNESADPMVSVTAPVLSKSPQFOLADVLNVRSDLEQVLTIEIEPE 1084

Qy 432 SVEEHVSQTSVPVQOSKODQ 451

Db 1085 AAKQEPQMSLEKSTKNVQ 1104

RESULT 8

hypoetical protein smc [imported] - *Streptococcus pneumoniae* (strain R6)

C:Species: *Streptococcus pneumoniae*

C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001

C:Accession: E98012

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Buggett, S.; Dehoff, B.S.; e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mchren, S. Y., P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.

A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: E98012

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1179 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99929.1; PID:g15458753; GSPDB:GN00174

C:Genetics:

A:Gene: smc

Query Match

Best Local Similarity 21.6%; Score 134.5; DB 2; Length 1179;

Matches 100; Conservative 80; Mismatches 178; Indels 105; Gaps 24;

Qy 5 KHQEQLSKPLSDSICGVYIKLEKSAFRLRNFPNNAOTLRKLSQNSPADERDALOEALNK 64

Db 674 KPELEQLOKEIADDE--ASLGSEALKTLQDQMAALTELEAIK--SGEQARIQEQ 727

Qy 65 CLNKWKILSDSLVEQFSKTRDIELISMFVAAGFLDPTLESANSLFWLADLSEKHMDH 124

Db 728 GLS-----LAVQQTQSQVBELET-LWKLOEE-EIDRLSEG-----DWQAD-KEKQES 772

Qy 125 L-----NPLVLPVETLKSDDKGKEREQADAKVKAFVOLVGDSESSSLIVAPVQLPL 176

Db 773 LATTASEKQMLEAEIEIRKSNKNAIOERYO-----NLQEEVAQARLLKTEL 818

Qy 177 VGEVTFPDPQSAERKGEISQLKSMLTITVAQERF-AIOFKMENAKRCVQOLRLSALVS 234

Db 819 QGQKR-YEVADIERLG-----KELDNINIEQEEIQRMLQEKVDNLEKVDTE-----LLS 866

Qy 235 TKCHSLGSGSTNGF--GFAKSL-LTRVENALVHLG-----IKLAPRAEAKTV 278

Db 867 QQAEESTQKTNTLQQLIRQFELDIEGQLDIALSLDQARQNEEMIRKQTRAEAK-- 924

Qy 279 EOEYAESVSSEGEPLSHMDTKHIERIPMASEQAOTVSOHLHAGNLG--ELGNLNNMRDL 336

Db 925 KEKVSERL-----RIHQQLTDQYQISYTALEKHAHELENNLAEQVQDLEKALRSL 977

Qy 337 AFHLREVSDFROSEPHSPISFLLEKAIKRWGLSLPELLREMSSEONGDA---LSTIFN 393

Db 978 GPVNLAIIDQY---EEVHNRLDPL--NSQRDILSANKLLLETTEENDEVKERFKSTFE 1032

Qy 394 AA-----GLNHLDOVLPEVSTPTVGISSPQEQAK 424

Db 1033 AIRESFVTFKQMFQGGQAD-LITTEGDLITAGVEISVQPPGK 1074

RESULT 9

118302

apsa protein - *Emeticella nidulans*

C:Species: *Emeticella nidulans*, *Aspergillus nidulans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-May-2000

C:Accession: T18302

R:Stelmann, R.; Sievers, N.; Galetzka, D.; Robertson, L.; Timberlake, W.E.; Fischer, M.O. Microbiol. 30, 831-842, 1998

A:Title: Increased nuclear traffic chaos in hyphae of *Aspergillus nidulans*: molecular

A:Reference number: Z18870; MUID:99140441; PMID:10094631

A:Accession: T18302

A:Status: preliminary; translated from GB/EMBL/DBJ



[illegible]

RESULT 12  
AE1643  
ATP-dependent dsDNA exonuclease SbcC homolog sbcC [imported] - *Listeria innocua* (strain  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 27-Nov-2001  
C:Accession: AE1643  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgeat, O.; Entlian, K.D.; Fsihi, H  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurupkat, G.; Madeno, E.; Maitournam, A.; Ma  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A.:Title: Comparative genomics of *Listeria* species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AE1643  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-1023 <GAP>  
A:Cross-references: GB:AL592022; PIDN:CAC96917.1; PID:916414173; GSPDB:GN00178  
A:Experimental source: strain Clip11262  
C:Genetics:  
C:Gene: sbcC

Query Match	5.4%	Score 126.5	DB 2	Length 1023
Best Local Similarity	20.3%	Pred. No. 3		
Matches 102	Conservative 82	Mismatches 192		Indels 127
				Gaps 21

```

Oy      8  IEO:SLPLSDSICGVYUJKELEKSAFRLRNENFVAOGALRKLSONPESADERDALOEJCUN 67
Db      284  IEEVAK-----LEAEKDFYGYIANKRIEAKRRASNRSDALCIRLKKQOLEFAVN 332
Oy      68  KWKILS---DLVEQFSKTTEDILIMFVAAOFLTLTTEASANSLEWLADSEK---- 120
Db      333  TEKOVAEVELITNOFSNAKKQKAL---AEQ---SAVLEANKRTLEQEEHEPKITEL 385
Oy      121  -----HW-----DHLNPLPYE-----TLKSDDDKQKREQADAKVAFQLV 159
Db      386  ETVYIQRRAELINEADRIETEKWKTEQOETIAELTOSMETRPLGTINAEI---ALLFAIN 442

```

```

OY 160 DSESSILYAPVLOLPEVTEFPDQSAKKEISOLKSMITLTYAORFPIO--FKME 217
Db 443 KRTTAEALIEK--NOELVNNRMKMDLNNKKOQREOETLQOULL-----EKLAIEIMKOE 495
OY 218 NAKRCVYOLDLSALV--STKCHSIGS---QSTNGFAKSLTTRYNALVHLSGIKLAP 271
Db 496 ESNIOQERBATLALHHEGDGACPCGSSVSHPOLAEVGEASLTLEE----- 542
OY 272 KAEAKTVEQEVYA---BSSVS--EGELPSHMDTKHTERTIPMASQAOIVSOHLHAGNISE 325
Db 543 -AKAKLHEKOLAIINEVKSISQIEMOLAEWADIADVDAVAEKILAE----- 588
OY 326 LGNNLNMMNRDLAFHLEAVSDYFQSPHSPISFLELKAIRMGYLSLPELLREMSSEONG 385
Db 589 -----NROLAKMLTIOIOLQOTNAOKENIGALE-----TLKKOKELEFTEKN 633
OY 386 DAL-----STINAGLNLHMDLOVLLPEVSPTYGIESPQPOAKPSVSDRSVEHNS 438
Db 634 TALOVESFHQEVOLSSKLSLTLEOALPADRLDKT-----FDKKNELSN--SITKHLE 685
OY 439 QTSPEVTOQSKODOKPOQSATSAL 461
Db 686 QAEQVDKTFREAEKETRLLESTL 708

```

RESULT 13  
S50914  
cell division control protein CDC37 - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: Cdc37 start control protein; protein YD9489.03; protein YDR166w  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 10-Feb-1995 #sequence\_revision 12-May-1995 #text\_change 29-Oct-1999  
C:Accession: S50914; B26372  
R:Olliver, K.; Harris, D.  
submitted to the EMBL Data Library, January 1994

A:Cross-references: EMBL:Z47813; NID:g642274; PIDN:CM87799.1; PID:g642277; MIPS:YDR168A; R.Ferguson, J., Ho, J.Y.; Peterson, T.A.; Reed, S.I.  
Nucleic Acids Res. 14, 6681-6697, 1986  
A>Title: Nucleotide sequence of the yeast cell division cycle start genes CDC28, CDC3  
A:Reference number: A93635; MUID:86312926; PMID:3018676  
A:Accession: B26372  
A:Molecule type: DNA  
A:Residues: 58-168, 'D', 170-506 <FER>  
A:Cross-references: EMBL:X04288; NID:g3492; PIDN:CA27836.1; PID:g3493  
A>Note: the authors translated the codon GAC for residue 112 as Ala  
C:Genetics:  
A:Gene: SGD:CDC37  
A:Cross-references: SGD:S0002575; MIPS:YDR168W  
A:Map position: 4R  
C:Keywords: cell cycle control

Query Match	5.4%	Score 126;	DB 2;	Length 506;
Best Local Similarity	19.1%;	Pred. No. 1.2;		
Matches 90;	Conservative 71;	Mismatches 161;	Indels 148;	Gaps 177;

```

OY 15 LSDOSICGVYLKLEKSARPLRNEFNNAQTALRLKLSQNPASDEADALQOEACLNKWKILSD 74
      ||||| : : : : |||
Db 13 LSDSDVEVHEHNVADKKSF-----IKW--OQ 36
      ||||| : : : :
OY 75 SLVEQFSKTTDIDELISFVAQAQFLDOTTLESANSLLEMLADLSEKHMHDHNPVLPVETL 134
      ||||| : : : : |||
Db 37 SIHEQRFRKNDIKNLETQVVMYSHLNKVRDRI-----LSNPESSITLDPATYKFLMA 90
      ||||| : : : : |||
OY 135 KSDDDKGERQADAKVAFAPOLYGDSESSIIYAPVQLPLVGEVYFFPDQSAERKEGI 194
      ||||| : : : : |||
Db 91 NEDKMEKSGENVDPPEIATYNEWEDLFE-----QLA-----KDLDEKGD 131
      ||||| : : : : |||
OY 195 SOLKSMILTTVAQEFALQFMEKNAKRCVYDLDRLSALVYKCHSLGSSQSTNPFASLL 254
      ||||| : : : : |||

```

Db 132 SKSPSLIRDAILKHKRAKIDSVTEAKK-----KIDELYKKNMAHISEDINTGPDSPFN 185  
OY 255 TRVENALVHLSGI---KLAPKAERKATVEOEVAESSVSE-----GEL 292  
Db 186 NKOGAKKLEATPPSALSSAAS-NILNKLAKSSVPQFFIDPKDDPKMLAKETBERGKI 244  
OY 293 PSHMDTKH---IERIPMASEQ-----AOTVSQHLHAG----- 321  
Db 245 SINEYSKQKFLLEHLPIISEQOKDALMKMAFEYQJHGDDKMTLOVHQSELMAYIKETI 304  
OY 322 -----NLSELGNLNNNNRDLAFHLLEVSDFRQSEPHSPHSIFLLEKAIMGYLSLP 373  
Db 305 DMKKIPLYLPMELSNVINM-----FFPKV--IFNKDKPMGKSEFL--RSVQEKFLHIQ 353  
OY 374 E-----LLREMSQNGDALSTIFNAGLNHLDOVLLPEVSTPYVGIESPQ 419  
Db 354 KRKKILQGEEMDESNAEGVETI---QLKSLDDSTELVNLPPDNKDPDE 399

## RESULT 14

T22235  
hypothetical protein F45G2.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22235  
R.Lindsay, S.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19535  
A:Accession: T22235  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1287 <MIL>  
A:Cross-references: EMBL:Z93382; PIDN:CA07612.1; GSPDB:GN00021; CESP:F45G2.3  
A:Experimental source: clone F45G2  
C:Genetics:  
A:Gene: CESP:F45G2.3  
A:Map position: 3  
A:introns: 59/2; 117/3; 153/3; 180/2; 233/2; 341/3; 393/1; 424/2; 551/3; 664/3; 734/3; 8

## Query Match

Best Local Similarity 5.4%; Score 125.5; DB 2; Length 1287;

Matches 90; Conservative 79; Mismatches 167; Indels 85; Gaps 18;

OY 7 QTEQLSKPLSDSICGVYKLEKSAPRPNNEFNVAQTALRKLSQNPADERALDEACL 66  
Db 259 EIEELNRNLKAESSENYENAKKASELERLR-----EKKEKDEKERKQMAENDEA-- 310  
OY 67 NKMKILSDLYEFSKTRTDIELISFVAQFLDITLESANSLDMLADLSEKHWDLN 126  
Db 311 NQHKQTETKRLQDITLQDDEFKVSQARKAQEQMNAELVDEVASFQKAKERADEQKKTL- 369  
OY 127 PVLVPVITLSDDDK-GKEHQADAKYKAFQVLGVGSESSILYAPVQLP-L-VGEVTFPD 184  
Db 370 ---VDLDSLDKDLAKERANNEHVKHKKLEGO-----LKATOTLTALKEKHEED 419  
OY 185 FQSAERKGEISOLKSMLTTFVAQERFAIOFKME-NAKRCVOTDRLSALVSTCHSLGSG 243  
Db 420 VQCKRRESIEGLK-----LKAQGDANLISKLOAMLRKCSRLEEDLEEDLEE----- 467  
OY 244 STNFGAKLLTRVENALVHLSGIKLAPAAKATVEQEAESSVSEGELPSHMDTKHIER 303  
Db 468 -----RKLRMKAERQFNEL-----RSEYEVLOEQMAEAS--GQLTAE--AHINK 507  
OY 304 IPMASQACTVSOHLHAGNLSE-----LGNLNNNR---DLAFHLLEVSVD 346  
Db 508 V--RAEVSNLRRDLQKRLNLEHAYISDLCNNQYATVNNLRNLSQOFSEFNCFEFLFL 565  
OY 347 YFRQSEPHSPIS-FLLEKAIMGYLSLPPELLREMSQNGDALST-----IFNAAGL 397  
Db 566 FFEFSFYDKPRHNFQIRKTIK---SPVTYVIFHLLISRESVPTVASLCITAYQTIIFNGASL 622  
OY 398 N 398

Db 623 N 623

## RESULT 15

A47297  
myosin heavy chain form B, nonmuscle - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 02-Feb-2001

C:Accession: A47297; A55441

R.Bhatia-Dey, N.; Adelstein, R.S.; David, I.B.

Proc. Natl. Acad. Sci. U.S.A. 90, 2856-2859, 1993

A:Title: Cloning of the cDNA encoding a myosin heavy chain B isoform of Xenopus nonmu

A:Reference number: A47297; MUID:93219383; PMID:8464900

A:Accession: A47297

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-1992 <BHA>

A:Cross-references: GB:I09740; NID:9214623; PIDN:AAA49915.1; PID:9214624

A:Experimental source: XTC cells

A:Note: sequence extracted from NCBI backbone (NCBIP:128722)

R.Kelley, C.A.; Oberman, F.; Yisraeli, J.K.; Adelstein, R.S.

J. Biol. Chem. 270, 1395-1401, 1995

A:Title: A Xenopus nonmuscle myosin heavy chain isoform is phosphorylated by cyclin-p

F:88-787/Domain: myosin motor domain homology <MMOT>

F:178-185/Region: nucleotide-binding motif A (P-loop)

## Query Match

Best Local Similarity 5.4%; Score 125.5; DB 2; Length 1992;

Matches 107; Conservative 78; Mismatches 189; Indels 131; Gaps 20;

OY 1 MPELSKQIDQLSKPLSDSICGVYKLEKSAPRPNNEFNVAQTALRKLSQNPADERALD 60  
Db 929 LAIKKQEMELIIDLDE-----IRMEEEERNVLYLNEKKKMTQVODLEE--QLDEEEA 980  
OY 61 LDQACLNK-----WKILSDSLY--EFSKTRTDIELISFVAQFLDITLESANSL 111  
Db 981 AQLQLEKVTAEAKIKKMEEDILVLDQNSKFLKELKLLLEERIAESTQSALAEERAKNL 1040  
OY 112 EWLADLSEKHWDLNLPVETLKSDDDKGKEHQADAKYKAFQVLGVGSESSILYAPV 171  
Db 1041 AKLKNQEMKISDL-----ERLKEEKTROELEKAKRK----- 1074  
OY 172 LDPLVGEVTFPDFOGSAERKGEISOLKSMLTTFVAQERFAIOFKMENAKRCVYQ----- 225  
Db 1075 ---LDGETTDFOQDLAEQLAQLEELQLL-----AKKEEQLALARGDEEVLQKNNITLK 1126  
OY 226 -LDRLSALVSTCHSLGSGSTNFGFA-----KSLTRVENALVHLSGIKLAPKA 273  
Db 1127 LVRELQALAEQLDELESEKASNNKAKKQKRLDSEELALKLTLEDTL-----DTAAQ 1181  
OY 274 EAKTV-EQEAESSV-EGELPSH-----MDTKHIERIPMASQACTVSOHLHAGN 322  
Db 1182 ELRTKREOEVAELRKSIIEETRNHEAQIOEMRQATALEEISEQLEQAKRFKVNLEKN 1241  
OY 323 LSELGNLNNNRDLAFHL-----LREVSDFRQ-----SEPHSPISFLLEKAIMGYLS 371  
Db 1242 QS-----LESNNKLATEVSSLOQMKASEYKKKKLGGVOELHAKV---LE----- 1285  
OY 372 LPELLREMSQNGDALSTIFNAAGLNHLDOVLLPEVSTPYVGIESPQTPQAKPSVSDPR 431  
Db 1286 -GRLRADVYKSSKQLONELENVS-----LLEAEKKKGIVL-----AKDVA 1326  
OY 432 SVEEHVSQTSPPVDTQSKQDQKPOSS 456  
Db 1327 SMESQDQTOELLQETROKLANOSS 1351

Search completed: March 2, 2003, 05:06:12  
Job time : 52 secs

---

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2003, 01:09:08 ; Search time 22 Seconds

(without alignments)  
872.888 Million cell updates/sec

Title: US-09-915-706A-2

Perfect score: 2343

Sequence: 1 MPLSKHQIEQLSKPLSDSI.....DTQSKQDQKPOSSATSALSW 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152	6.5	2245	1	MYSJ_DICDI
2	126	5.4	506	1	CC37_YEAST
3	125	5.3	793	1	REGA_DICDI
4	124.5	5.3	1940	1	MYS3_HUMAN
5	124.5	5.3	1940	1	MYS3_RAT
6	124	5.3	866	1	MYSF_SCHMA
7	123.5	5.3	867	1	DPOI_RICPR
8	123	5.2	1935	1	MYS7_HUMAN
9	123	5.2	1935	1	MYS7_RAT
10	123	5.2	2022	1	ANT1_ONCVO
11	122.5	5.2	2469	1	TEGU_HSVSA
12	122	5.2	1934	1	MYS7_MESAU
13	120.5	5.1	999	1	OKRP_HUMAN
14	120.5	5.1	1938	1	MYS4_RABIT
15	120	5.1	650	1	M10_CAEEL
16	119.5	5.1	1066	1	KL61_DROME
17	119	5.1	1935	1	MYS7_PIG
18	119	5.1	2104	1	MYS3_SCHPO
19	118.5	5.1	1920	1	PCNT_MOUSE
20	118.5	5.1	1937	1	MYS8_HUMAN
21	117.5	5.0	539	1	YAH2_YEAST
22	117.5	5.0	1427	1	REST_HUMAN
23	117.5	5.0	2116	1	MYS2_DICDI
24	117	5.0	1679	1	YIO9_YEAST
25	116.5	5.0	4385	1	YF73_CAEEL
26	116	5.0	1616	1	P200_MYCE
27	116	5.0	1966	1	MYSB_CAEEL
28	115.5	4.9	2468	1	MABP_HUMAN
29	115.5	4.9	3660	1	DMD_CHICK
30	115	4.9	1312	1	RA50_YEAST
31	114.5	4.9	955	1	KIRL_LEICH
32	114.5	4.9	2230	1	GOG4_HUMAN
33	114.5	4.9	5327	1	ACE7_MOUSE

34	114	4.9	866	1	MYSF_SCHMA	005870 schistosoma
35	114	4.9	1940	1	MYS3_CHICK	P02565 gallus gall
36	114	4.9	3672	1	LML2_CAEEL	021313 caenorhabd
37	113.5	4.8	671	1	CHEA_BACSU	P29072 bacillus su
38	113	4.8	524	1	KZC4_MOUSE	P07744 mus musculi
39	113	4.8	707	1	YJ9C_YEAST	P47166 saccharomyc
40	112	4.8	804	1	MEA6_HUMAN	015320 homo sapien
41	112	4.8	1939	1	RECN_HETPJ	P13533 homo sapien
42	111.5	4.8	522	1	YR34_HUMAN	092180 helicobacte
43	111.5	4.8	733	1	YR34_HUMAN	092180 helicobacte
44	111.5	4.8	1939	1	MYS1_HUMAN	092180 helicobacte
45	110.5	4.7	1252	1	RPOB_CHLPP	092990 chlamydia p

## ALIGNMENTS

RESULT 1	ID	MYSJ_DICDI	STANDARD:	PRT:	2245 AA.
AC	P54697:				
DT	01-OCT-1996 (rel. 34, created)				
DT	01-OCT-1996 (rel. 34, last sequence update)				
DT	16-OCT-2001 (rel. 40, last annotation update)				
DE	Myosin II heavy chain.				
GN	MYOJ.				
OS	Dicystostellium discoideum (Slime mold).				
OC	Eukaryota; Mycetozoa; Dicystostellida; Dicystostellium.				
OX	NCBI_TaxID=44689;				
ON	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=AX3:				
RX	MEDLINE=96215148; PubMed=8636147;				
RA	Hammer J.A. III, Jung G.;				
RT	"The sequence of the dicystostellium myo J heavy chain gene predicts a				
RT	novel, dimeric, unconventional myosin with a heavy chain molecular				
RT	mass of 258 kDa."				
RL	J. Biol. Chem. 271:7120-7127(1996).				
RL	[2]				
RP	SEQUENCE OF 1-1021 FROM N.A.				
RX	MEDLINE=97039016; PubMed=8884597;				
RA	Peterson M.D., Urioste A.S., Titus M.A.;				
RT	"Dicystostellium discoideum myof: a member of a broadly defined myosin				
RT	V class or a class XI unconventional myosin?";				
RL	J. Muscle Res. Cell Motil. 17:411-424(1996).				
RL	[3]				
RP	SEQUENCE OF 182-298 FROM N.A.				
RX	MEDLINE=95023928; PubMed=7937787;				
RA	Titus M.A., Kuspa A., Loomis W.F.;				
RT	"Discovery of myosin genes by physical mapping in Dicystostellium.;"				
RL	Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).				
CC	- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE				
CC	ACTIVITY THAT IS ACTIVATED BY ACTIN.				
CC	- SUBUNIT: HOMODIMER.				
CC	- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.				
CC	- SIMILARITY: CONTAINS 3 IO DOMAINS.				
CC	- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
DR	EMBL: U32409; AAA85186.1; -				
DR	EMBL: L35322; AAA79858.1; -				
DR	HSP: P08799; 1MND.				
DR	DicystDb: DP01095; myoj.				
DR	InterPro: IPR002710; DIL.				
DR	InterPro: IPR000048; IO_region.				
DR	InterPro: IPR004009; Myosin_N.				

DR InterPro: IPR001609; myosin\_head.  
 DR Pfam: PF00063; myosin\_head; 2.  
 DR Pfam: PF00612; IQ; 6.  
 DR Pfam: PF01843; DIL; 1.  
 DR Pfam: PF02736; Myosin\_N; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR ProDom: PD000355; myosin\_head; 2.  
 DR SMART: SM003376; DIL; 1.  
 DR SMART: SM00015; IQ; 3.  
 DR SMART: SM00242; MYSC; 1.  
 DR PROSITE: PS50096; IQ; 3.  
 KM Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;  
 KM Coiled coil.  
 FT DOMAIN 1 809 MYOSIN HEAD-LIKE.  
 FT DOMAIN 824 851 IQ 1.  
 FT DOMAIN 872 901 IQ 2.  
 FT DOMAIN 943 972 IQ 3.  
 FT DOMAIN 973 1812 COILED COIL (POTENTIAL).  
 FT DOMAIN 1813 2245 TAIL.  
 FT NE\_BIND 174 181 ATP (POTENTIAL).  
 FT DOMAIN 669 749 ACTIN-BINDING.  
 FT CONFLICT 191 191 L -> F (IN REF. 2).  
 FT CONFLICT 284 284 A -> T (IN REF. 2).  
 FT CONFLICT 291 291 G -> R (IN REF. 2).  
 FT CONFLICT 332 347 MSGCFEIEGVSDEEH -> IEMFELKVRMKS (IN REF. 2).  
 FT CONFLICT 550 550 N -> K (IN REF. 2).  
 FT CONFLICT 865 866 HH -> QQ (IN REF. 2).  
 SO SEQUENCE 2245 AA; 258478 MW; 615E5EF1DA845BE CAC64;

Query Match 6.5%; Score 152; DB 1; Length 2245;  
 Best Local Similarity 23.6%; Pred. No. 0.11;  
 Matches 81; Conservative 57; Mismatches 143; Indels 62; Gaps 13;

QY 23 VYLKLSARPLRNENNAQVLRKLSQNPADALQACLNKKKILSDSLYEFSK 82  
 DB 964 IQLAEKRSRTVOEQNNKLOEKLEELQWRITSEAKRKQ--LEDQKVSDDTISLS 1020  
 QY 83 TTRDIEL-ISMFWAQCFLDPTLESANSL-EMLDLSEK---HMDHNPVLPVETLK 135  
 DB 1021 NNHLEQLQLSEIQKQVDELNSNSQQLSECKSLKEQTQQLDHSKLNKKLEKLS 1080  
 QY 136 SDDDKGERQADAKKAFQVLGVDSSESSILYAPVQLPLVGEVF---PDFQSAER- 190  
 DB 1081 QHDSIEKLOQFNETEQLOQFKQSESELSKLSKTTQ---QLDANKQEFDLQOERD 1135  
 QY 191 -----KGEISQKSMLTVAQERFAIQFKNEMAKRCVQLDLRLSLVSKHSLGSS 244  
 DB 1136 TDTNTNQLLEIQOLKK-ANSTLEEDYFSLGIRNLEKQVLELRDENQILKERLSD 1194  
 QY 245 TNF--GFA-----KSLLTREVNALVHLSGIKLAAPKAER-----TVE 279  
 DB 1195 SDFQSGAALKEQQLQEQVQSEQLIKLSEKLGSEEEKQINQLELTLDRKSLQIQ 1254  
 QY 280 QEVAEVS-----EGELPSHMDTK-----HIERIPMASEQQ 312  
 DB 1255 LQLTEOSNEKIKKLKLEEYQDEKQLOEQLEKRIKOSKOSVE 1297

RESULT 2  
 CC37\_YEAST  
 ID CC37\_YEAST STANDARD; PRT; 506 AA.  
 AC P06101; Q04132;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE Hsp90 co-chaperone Cdc37 (Hsp90 chaperone protein kinase-targeting subunit) (Cell division control protein 37).  
 GN CDC37 OR SMO1 OR YDR168W OR YD9489.03.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxId=4932;

RN [1]  
 RP SEQUENCE OF 58-506 FROM N.A.  
 RX MEDLINE=66312926; PubMed=3018676;  
 RA Ferguson J., Ho J.-Y., Peterson T.A., Reed S.I.;  
 RT "Nucleotide sequence of the yeast cell division cycle start genes CDC28, CDC36, CDC37, and CDC39, and a structural analysis of the predicted products.";  
 RL Nucleic Acids Res. 14:6681-6697(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=5288C / AB972;  
 RA Oliver K., Harris D., Barrell B.G., Rajandream M.A.;  
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP INTERACTION WITH CDC28.  
 RX PubMed=7753858;  
 RA Gerber M.R., Farrell A., Deshaies R.J., Herskowitz I., Morgan D.O.;  
 RT "Cdc37 is required for association of the protein kinase Cdc28 with G1 and mitotic cyclins.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:4651-4655(1995).  
 RN [4]  
 RP INTERACTION WITH STELL.  
 RX PubMed=10664467;  
 RA Abbas-Tekki T., Donze O., Picard D.;  
 RT "The molecular chaperone Cdc37 is required for Ste11 function and pheromone-induced cell cycle arrest.";  
 RL FEBS Lett. 467:111-116(2000).  
 RN [5]  
 RP INTERACTION WITH CDC28 AND CAK1.  
 RX PubMed=10629030;  
 RA Farrell A., Morgan D.O.;  
 RT "Cdc37 promotes the stability of protein kinases Cdc28 and Cak1.";  
 RL Mol. Cell. Biol. 20:749-754(2000).  
 CC -1- FUNCTION: With Hsp90 it forms a complex that binds to several kinases, resulting in stabilization and promotion of their activity.  
 CC -1- SUBUNIT: Forms a complex with Hsp90. Interacts with CDC28, CAK1 and STELL.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO THE CDC37 FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X04288; CAA27836.1; -;  
 DR EMBL: Z47813; CAA87799.1; -;  
 DR PIR: B26372; B26372.  
 DR SGD: S0002575; CDC37.  
 DR InterPro: IPR004918; Cdc37.  
 DR Pfam: PF03234; Cdc37; 1.  
 KW Chaperone; Cell division; Cell cycle.  
 FT CONFLICT 169 169 A -> D (IN REF. 1).  
 SO SEQUENCE 506 AA; 58385 MW; 0DF0C923158A2526 CRC64;

Query Match 5.4%; Score 126; DB 1; Length 506;  
 Best Local Similarity 19.1%; Pred. No. 0.64;  
 Matches 90; Conservative 71; Mismatches 161; Indels 148; Gaps 17;

QY 15 LSDPSICGVYIKLEKSAFRLRNENNAQVLRKLSQNPADALQACLNKKKILSD 74  
 DB 13 LSDSDVEVPPNDKKE-----IKKK--QQ 36  
 QY 75 SLVEQFSKTRDIELISMFWAQCFLDPTLESANSLLEWLADLSEKHWDLNPLVETL 134  
 DB 37 SIHQREKRNQDILKNETQVDMYSHLNKRVDR-----LSNPESSLTDLPAVTFPLNA 90  
 QY 135 KSDDDKGERQADAKKAFQVLGVDSSESSILYAPVQLPLVGEVTFPFQSAERGET 194



```

Db 91 NEDKMEKSGENVPEIATYNEKEDLFE-----OLA-----KDLDEKGD 131
QY 195 SOLKSLTTTVAQERFAIQFKMENAKRCYTQDLRLSALVSTKCHSLGOSOTNGFAKSL 254
Db 132 SKSPSLIRAILKHKAKISVYEAKK-----KLDLKEKNAHLSSEDIHTGDSSEFM 185
QY 255 TRVENALVLSGI---KLAPKAKEKTEQEVASSVSE-----GEL 292
Db 186 NKQKGAQKLEATPSEALSAAES-NILNKLAKSSVPQFTIDKDDPKLAKETEERFKI 244
QY 293 PSMIDTKH-----IERIPMASEQ-----AQTVSOHLHAG-----321
Db 245 SINEYKSKQKFLLEHLPIISEQCKDALMKAFEQQLHGDKMTLQYIHQSELMAYIKEIY 304
QY 322 -----NLSLGNLNNNNRDLAFHLREVSDFROSEPHSPISFLLEKAIHWGLSLP 373
Db 305 DMKKIYLPNEMELSNVINM-----FPEKV--TFNKDKPMGSESL--RSVOEKLHIQ 353
QY 374 E-----LLREMSSEONGDALSTIFNAGLNHLDOVLLPEVSTPVGIESPO 419
Db 354 KRSLIOGEEMDESNAEGVETI-----QLKSLDSTLEVLNLPDPSKDP 399

RESULT 3
REGA.DICDI STANDARD: PRT: 793 AA.
AC 023917:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3',5'-cyclic-nucleotide phosphodiesterase rega (EC 3.1.4.17) (PDBase
DE rega).
GN REGA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
ON NCBI_TaxID=44689;
RX SEQUENCE FROM N.A.
RC STRAIN-AX4:
RX MEDLINE-97140317; PubMed-8986798;
RX Shaulsky G., Escalante R., Loomis W.F.;
RT "Developmental signal transduction pathways uncovered by genetic
RT suppressors.";
RT Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265(1996).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AX2:
RA Thomson P.A., Traynor D., Cavet G., Chang W.T., Harwood A.J.,
RA Kay R.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DDA databases.
CC -1- FUNCTION: MORPHOLOGICAL SUPPRESSOR OF TAGB.
CC -1- CATALYTIC ACTIVITY: Nucleoside 3',5'-cyclic phosphate + H(2)O =
CC nucleoside 5'-phosphate.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED AT LOW LEVELS IN VEGETATIVE CELLS
CC AND AT HIGH LEVELS IN PRESPORE AND PRESTALK CELLS DURING
CC DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE PHOSPHODIESTERASE
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC -----
CC EMBL: U60170; AAB03508.1;
CC EMBL: A005398; CAA0513.1;
CC DICTYDB: DD02055; REGA.
CC InterPro: IPR003607; ME_Plase_HDC.
CC InterPro: IPR002073; PDase.

```

```

DR InterPro: IPR001789; Response_reg.
DR Pfam: PF000733; PDase.1.
DR ProDom: PD000039; Response_reg.1.
DR SMART: SM00471; HDC.1.
DR SMART: SM00448; REC.1.
DR PROSITE: PS00126; PDASE_1; 1.
DR PROSITE: PS00110; RESPONSE_REGULATORY.1.
KW Hydrolase; CAMP; cGMP.
FT DOMAIN 161 280 RESPONSE REGULATORY.
FT DOMAIN 18 28 POLY-SER.
FT DOMAIN 52 69 POLY-ASN.
FT DOMAIN 87 96 POLY-THR.
FT DOMAIN 100 121 POLY-ASN.
FT DOMAIN 166 169 POLY-ASP.
FT DOMAIN 770 776 POLY-SER.
SQ SEQUENCE 793 AA; 91175 MW; 6E065A620FBC27 CRC64;

Query Match 5.3%; Score 125; DB 1; Length 793;
Best Local Similarity 19.0%; Pred. No.1.3;
Matches 107; Conservative 88; Mismatches 183; Indels 184; Gaps 27;

QY 1 MPLSKHQIQBLQSKPLS-----DSDICGVYKLE---KSAERPLRNEF-----39
Db 310 MELKEHEIEELTKKVSMSISKAMESPLVSTRNIEELKQSSMSHYSEIKEKLSI 369
QY 40 -----NVAQTALRLKSQNPASDERDALQACLNKWKILSDLYQFSKTTDIELISW 92
Db 370 LKELGSSNTRYRPFELKINKDSVD-----PVKSFVSEFSTT-----408
QY 93 FVAAGFLDTLTLSAANSLEMLADLSEKHMDLNPVLPVETLKSDD---GKEREQADA 149
Db 409 -----SRNST-----PRFQTTYRDRKEVYKWE-----434
QY 150 KVAFQVLQVDSSESSILYAVLQDLPLVGEVTFDFQSAE-RKGEISQKSLMTTVADE 208
Db 435 -----FDVFKYSDD-----LMPILVDM-FENFQLPRIKFIPIEKLQRFIMTNALY 480
QY 209 R-----FALQFKME-----NAKRCVQDLRLSALVSTKCHSLGOSOTNGFAK 251
Db 481 RKNRRYHNHFADVDTQVYVTFLLSFNAAOYLNLHDLIFALLISCMDLHNPENNTFOY 540
QY 252 SLTR-----VENALVHLSGIRKLAPKAKEVSE-----OEVAESSV---EGE 291
Db 541 NAQTELSLEYNDISVLENHNAMLT-FKILRNSCNILEGNEQYKELRSVQILLATD 599
QY 292 LPSHMD-----TKHIERIPMASDAQTVSQHLHAGNLSLGNLNNNRDLAFHL-----L 341
Db 600 MGNHFEHTNKFQHNLNLPF--DRNKKEKQMLNLFILKCGDISNIR--PMHINFEWSL 655
QY 342 REVSDYFROSEPHS-----PISFLEKA-----IRNGYL-----SLPELLREMA--380
Db 656 RVSDDEFQOQSHYETIGYRPTPMDTKTTRARIADDFIDFVASPLFQSAKFLKSQFL 715
QY 381 -----SEONGDALSTIFNAGLNHLDOVLLPEVSTPVGIESPOTPQAPVSDPRSYE 456
Db 716 LKVISKNREMWQAQYMEQKGEKCNDD---LQFMEDPTILVKS-KLPIKIDEENRDKVSS 771
QY 435 EHVQSOTSPVDTQSKODQKPOSS 456
Db 772 SSSSTAPLSTSSNNETSS 793

RESULT 4
MYH3_HUMAN STANDARD: PRT: 1940 AA.
ID MYH3_HUMAN
AC P11055; O15492:
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, fast skeletal muscle, embryonic (muscle embryonic
DE myosin heavy chain) (SMCE).
GN MYH3.

```



ID	MYH3_RAT	STANDARD:	PRT:	1940 AA.
AC	P12847;			
DT	01-OCT-1989 (Rel. 12, Created)			
DT	01-OCT-1989 (Rel. 12, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Myosin heavy chain, fast skeletal muscle, embryonic.			
CN	MYH3.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=87060988; PubMed=3783701;			
RA	Strehler E.E., Strehler-Page M.-A., Periard J.C., Perlansamy M.,			
RA	Nadal-Giard B.;			
RT	"Complete nucleotide and encoded amino acid sequence of a mammalian			
RT	myosin heavy chain gene. Evidence against intron-dependent evolution			
RT	of the rod."			
RL	J. Mol. Biol. 190:291-317(1986).			
CC	-1 FUNCTION: MUSCLE CONTRACTION.			
CC	-1 SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2			
CC	HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)			
CC	AND 2 REGULATOR LIGHT CHAIN SUBUNITS (MLC-2).			
CC	-1 SUBCELLULAR LOCATION: Thick filaments of the myofibrils.			
CC	-1 DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING			
CC	CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,			
CC	CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.			
CC	-1 PIM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY			
CC	ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.			
CC	-1 MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT			
CC	MEROMYOXIN (LM) AND 1 HEAVY MEROMYOXIN (HM). IT CAN LATER BE			
CC	SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED			
CC	SUBFRAGMENT (S2).			
CC	-1 SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.			
CC	-1 SIMILARITY: CONTAINS 1 IQ DOMAIN.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isdb-sib.ch/announce/">http://www.isdb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
CC	-----			
DR	EMBL; X04267; CAA27817.1; -.			
DR	PIR; A24922; A24922.			
DR	HSSP; P13538; 2MYS.			
DR	InterPro: IPR0000048; IO_region.			
DR	InterPro: IPR0004009; Myosin_N.			
DR	InterPro: IPR002928; Myosin_tail.			
DR	InterPro: IPR002017; Spectrin.			
DR	InterPro: IPR001609; myosin_head.			
DR	Pfam; PF00063; myosin_head_1.			
DR	Pfam; PF00612; IO; 2.			
DR	Pfam; PF01576; Myosin_tail_1.			
DR	Pfam; PF02735; Myosin_N_1.			
DR	PRINTS; PR00193; MYOSTINHEAVY.			
DR	Prodorm; PD000355; myosin_head; 1.			
DR	SMART; SM00015; IO; 1.			
DR	SMART; SM00242; MYSC; 1.			
DR	PROSITE; PS50096; IO; 1.			
KW	Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;			
KW	Calmodulin-binding; ATP-binding; Methylation; Alkylation;			
KW	Multigene family.			
KW	DOMAIN	781		MYOSIN HEAD-LIKE.
FT	DOMAIN	1		IO.
FT	DOMAIN	782	811	COILED COIL (POTENTIAL).
FT	NP_BIND	840	1933	ATP (POTENTIAL).
FT	DOMAIN	179	186	ACTIN-BINDING.
FT	DOMAIN	656	678	ACTIN-BINDING.
FT	MOD_RES	758	772	METHYLATION (TR1-) (POTENTIAL).
FT	MOD_RES	130	130	

FT	MOD_RES	696	696	ALTYLATION (SH-1).
FT	MOD_RES	706	706	ALTYLATION (SH-2).
SEQ	SEQUENCE	1940 AA:	223857 MW: B5D546A59662A696 CRC64:	
Query Match				
Best local Similarity		20.9%	Pred. No. 4.9:	
Matches 101:		Conservative 90:	Mismatches 194:	Gaps 99:
QY	7	QIEQLSPKLSDDSGCYLLKLSAER---	PLRNEFNVAQTALRKLSONPSADERDALQ	62
DQ	919	KIKFEVTERAEDEEINELNAETAKRKLEDECCSELKIDIDLETLTAKVEKEKHATE----		973
QY	63	EACLNKKKILSDSLY---	EOPFSKTRTDIELISFVAAQFLDD--	TLTLESAANSLEMLADL 117
DQ	974	---NKKYNLTLEELAGLDETIAKLTREKKALQ--	EAAHQOTLDDLOAEEDKVNLSKLKSK	1027
QY	118	SEKIMHDILNPILPVE-TLKSDDDKGKEERADAKV--	KAFVOLGSEESSILYAPVLQ	174
DQ	1028	LEQYDDLESSLDEEKKLRVLEERKKRLEEGDILKLAQESITLLENDKQO-----		1078
QY	175	PLVGEVTFPFDSAEKRGELISOLKSMLTTPVAOER--	PAIDFKMENAKRCVTOL-DRLS	230
DQ	1077	-----LBERLKKKQFEVSQLOQ-----	KVEDQGLTSLQLOKKIKELQARLELEEELE	1122
QY	231	ALVSTKCHSLGOSTNGFPAKSLTLTRYENA-LVHLSGILKAPKAENK--	TVEQEVAAESSV	287
DQ	1125	AERATPRAKTEKORSRDYARELEELSERLEEGAGVTSQIEELKKREAEFLKRLRDEEATL		1184
QY	288	SEGELPSIMDTKHIERIPMASEDAQVTSQIHLANGNISSELGNLNNMRDL----	AFHILR	342
DQ	1185	QHEATVATLRRKRD--	SAEELAE-----QIDNLORVAKOLEKEKSEFKL--	1227
QY	343	EVSDYFPOSEPHSPISFLKEAIRMGYLSLPELLREMSQONDALSTIFNAAGLNHLDQ		402
DQ	1228	EIDDLSSSVESVSKRANLEKICR-----	TLEQDLSE-----	ARGKKEEFO 1266
QY	403	VLLPEVSTPTVYGIESPQTPQAKPSVSDPSRSEEVHSQTSFVDTQSKODKPO---	SSAT	458
DQ	1269	RLSELFTQKSRLOQT-BAGELSLQLEEKESIVQSLSRKAQAFQOITELKRLQLEEBNKA		1327
QY	459	SALS 462		
DQ	1328	NALA 1331		
RESULT 6				
MTSP_SCMA	ID	MTSP_SCMA	STANDARD:	PRT: 866 AA.
AC	P06198:			
DT	01-JUN-1988	(Rel. 06, Created)		
DT	01-JUN-1994	(Rel. 29, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	Paramyosin.			
OC	Schistosoma mansoni (Blood fluke).			
OC	Eukaryota: Metazoa: Platyhelminthes: Trematoda: Digenea: Strigeiida:			
OC	Schistosomatoidea: Schistosomatidae: Schistosoma.			
OX	NCBI_TaxID-6183:			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE-91270282: PubMed-2052029;			
RA	Laclette J.P., Landa A., Arcos L., Williams K., Davis A.E.,			
RA	Shenmaker C.B.;			
RT	"Paramyosin in the Schistosoma mansoni (Trematoda) homologue of			
RT	antigen B from Taenia solium (Cestoda)."			
RL	Mol. Biochem. Parasitol. 44:287-296(1991).			
RN	[2]			
RP	SEQUENCE OF 303-742 FROM N.A.			
RA	MEDLINE-87018840: PubMed-3094144;			
RA	Lanar D.E., Pearce E.J., James S.L., Sher A.;			
RT	"Identification of paramyosin as schistosome antigen recognized by			
RT	intradermally vaccinated mice."			
RL	Science 234:593-596(1986).			
CC	-1- FUNCTION: PARAMYOSIN IS A MAJOR STRUCTURAL COMPONENT OF			

CC MANY THICK FILAMENTS ISOLATED FROM INVERTEBRATE MUSCLES.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC -1- SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M35499; AAA29915.1; -;  
 CC DR EMBL: M4163; AAA29914.1; -;  
 CC DR PIR: A25993; A25993.  
 CC DR InterPro: IPR002928; Myosin tail.  
 CC Pfam: PF01576; Myosin\_tail; 1.  
 CC Colled coil; Muscle protein; Thick filament; Myosin.  
 CC KW DOMAIN 1 22  
 CC FT DOMAIN 23 839  
 CC FT DOMAIN 840 866  
 CC FT DISULFID 750 750  
 CC FT CONFLICT 423 428  
 CC FT CONFLICT 430 431  
 CC FT CONFLICT 637 637  
 CC FT CONFLICT 639 639  
 CC FT CONFLICT 691 691  
 CC FT CONFLICT 720 720  
 CC SQ SEQUENCE 866 AA; 100387 MW; 42FA56EF8176AE0 CRC64;  
 Query Match 5.3%; Score 124; DB 1; Length 866;  
 Best Local Similarity 21.3%; Pred. No. 1.8;  
 Matches 103; Conservative 82; Mismatches 182; Indels 116; Gaps 21;

QY 7 QIBLSPKPLSDSICGYVLEKSAFRLRNEFVAQTALFKLSQNSADRDALQEQCL 66  
 DB 122 EENMLQKQK-----KAEDKSHLMEVDNVLQDGLAKAKQSAESKLEGLDSOL 172  
 QY 67 NKMKILSDLSLEYQFSK-----TTTRDIEL-TSMFVAQFLDTEESANSLEWLAD 116  
 DB 173 NRLKSLTDDLOQLTELNNAKSRLTSENFELHTNDEQAOL---NYSKAKSSLEQYVD 229  
 QY 117 LSEKHWHLNPVLEVEYLKSDDKGKEREQADAKVAFQVLGD-----SEESSIL 167  
 DB 230 -----DKRSLLDDEAKNRFNLQAOITS-LQMDYDNLQAKYDESEESASNL 273  
 QY 168 YAPVLOLPLVGEVTFEPDQSAERKGEISQKSMLTTPVAOERFALIOFMENAKKCVOLD 227  
 DB 274 RS-----QVSKFNADIALAKSKFE--RELMSKTEEFEEEMKRKFTMR-----ITELE 317  
 QY 228 RLKSLVSTKCHSLGSGSTNFGF-AKSLTLRVENALVHLGKIKLPAKAEKTVDEQVAVSS 286  
 DB 318 DTAERERLKAVSLKELKTKLLEIKDLQSEIES--LSLENELIRRAKA-----AESL 368  
 QY 287 VSESELPDSHMDTKIETIRPMASEQAQTVSOHLHAGNL-SELGINNN-MNRDLAFHLREV 344  
 DB 369 AS--DLQRRVDLTIETVNTLSQNSQSENLRLKSLVNDLTDRKNLLEIRE-----NROM 421  
 QY 345 SDYFRQSPHPISFLLEKALRMGYLSPEL-LREMMSENGALSTIFNA----- 394  
 DB 422 NDQYKE-----LKSSLRDANRRRTDLEALRSOLEARDNLASLHDAEALHDM 471  
 QY 395 -----AGLNHLDOVLLPEVSTPTVGIESPOTQAPKPSVDPSPSEHVHVSQTSVPDTS 447  
 DB 472 QKYASQAALNHLKSEMQRLREDEBELSLRK-----STTTIEELVTTITEMEYKY 524  
 QY 448 KOD 450  
 DB 525 KSE 527

RESULT 7

DP01\_RICPR  
 ID DP01\_RICPR STANDARD; PRT; 867 AA.  
 AC 005949;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA polymerase I (EC 2.7.7.7) (POL I).  
 GN POLA OR RP776.  
 OS Rickettsia prowazekii.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;  
 CC Rickettsiaceae; Rickettsiaceae; Rickettsia.  
 CC NCBI\_TaxID=782;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Medrid E;  
 RC MEDLINE=97419517; PubMed=9274032;  
 RX Anderson J.O., Anderson S.G.E.;  
 RA "genomic rearrangements during evolution of the obligate  
 RT intracellular parasite Rickettsia prowazekii as inferred from an  
 RL analysis of 52015 bp nucleotide sequence";  
 RN Microbiology 143:2783-2795(1997).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Medrid E;  
 RX MEDLINE=99039499; PubMed=9823893;  
 RA Anderson S.G.E., Zomorodipour A., Andersson J.O.,  
 RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,  
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;  
 RT "The genome sequence of Rickettsia prowazekii and the origin of  
 RT mitochondria";  
 RL Nature 396:133-140(1998).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Medrid E. and B;  
 RC MEDLINE=99416441; PubMed=10486973;  
 RX Andersson J.O., Andersson S.G.E.;  
 RT "genome degradation is an ongoing process in Rickettsia";  
 RL Mol. Biol. Evol. 16:1178-1191(1999).  
 CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE  
 CC EXHIBITS 3' TO 5' AND 5' TO 3' EXONUCLEASE ACTIVITY (BY  
 CC SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate -> N diphosphate  
 CC + [DNA](N).  
 CC -1- SUBUNIT: SINGLE-CHAIN MONOMER WITH MULTIPLE FUNCTIONS (BY  
 CC SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-A FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Y11784; CAAT72476.1; -;  
 CC DR EMBL: AJ235273; CAAT5203.1; -;  
 CC DR EMBL: AJ238755; CAB56085.1; -;  
 CC DR EMBL: AJ238756; CAB56089.1; -;  
 CC DR HSSP: P00582; 1KFS.  
 CC DR InterPro: IPR002421; 5\_3\_exonuclease.  
 CC DR InterPro: IPR001098; DNA\_pol.  
 CC DR InterPro: IPR002298; DNA\_pol.  
 CC DR InterPro: IPR000513; Exo\_N.I.  
 CC DR InterPro: IPR003584; HHH\_2.  
 CC Pfam: PF004476; DNA\_pol\_A; 1.  
 CC DR Pfam: PF01367; 5\_3\_exonuclease; 1.  
 CC DR Pfam: PF02739; 5\_3\_exonuc\_N; 1.  
 CC DR PRINTS: PR00868; DNAPOLI.  
 CC DR SMART: SM00475; 53EXOC; 1.  
 CC DR SMART: SM00379; Hhh2; 1.  
 CC DR SMART: SM00482; POLAC; 1.  
 CC DR TIGRfams: TIGR00593; pole; 1.

DR PROSITE: PS00447; DNA\_POLYMERASE\_A: 1.  
KW Transferase: DNA-directed DNA polymerase; DNA replication; DNA repair;  
KW Hydrolyase: Exonuclease; DNA-binding; Complete proteome.  
FT DOMAIN 1 283 EXONUCLEASE  
SQ SEQUENCE 867 AA; 98651 MW; 557194B38FB69BA2 CRC64;  
Query Match 5.3%; Score 123.5; DB 1; Length 867;  
Best Local Similarity 21.2%; Pred. No. 1.9;  
Matches 90; Conservative 60; Mismatches 172; Indels 103; Gaps 17;  
QY 18 DSGCYVLLLEKSAFRLPNEFVAQTALRLKLSQNPADRALQACLKMTLSDSL 77  
DB 186 DNIQGVPSIGPKTA-SLLTFKFSVENIFNSLDQISSIKORKTLQNA----- 231  
QY 78 EQSKTTRDLELTSWFVAQFLDFT--LESANSLLEMLADSEKHMHLNPLPYETLK 135  
DB 232 -----REKALISMKLIG---LDSNVDDDFDNLNLMKSPSPNKK---LTGFLQEXGFK 277  
QY 136 SDDDKGERQADAKKAFQOL-VGDESESSILYAPVLOPLVGEVTFPFGQAEKKEI 194  
DB 278 S-----LKKRVENLFDIKINDHE-----IYDNKYTEAK-EI 308  
QY 195 SGLKSLMTTVAQER-----FAIQKRENAKRCVTQLDRLSALVSTKCHLSQSTNFG 248  
DB 309 SNASELNPAKEAEERIGIGIYLQOKGEN-RALLISLQNSYIITNNNNYNIKNNDW 367  
QY 249 FAKSLTFRVEN-----ALVHLSGIKLAPKAEATVER-EVAESSVSSEGLPSHMDTK 299  
DB 368 FSHITLNLTKNSIKKITYSKLHLKFPYANQSHQITAELELMQYALSGVLQOKMLFK 427  
QY 300 HIERIPMAEQAFVSO--HLHAGNLSLGNLNNMRDLAFHLLREVSDFRQSEPHSP 357  
DB 428 TLTKNDINNSAIVINFLSYKQTLLEL-----QKNKAFRLKYRED-----LPT 472  
QY 358 SPLEKAIKNGYLSPELLREMSSEONGAL-----STTFNAAGINHLDOVLLPE 407  
DB 473 CFILDMKGVKIVDANYINRLSDEFGTEILKIEEIFALSGTFNIGSQKQGLLEFK 532  
QY 408 VSRPT 412  
DB 533 MQLPS 537  
RESULT 8  
MYH7 HUMAN  
ID MYH7 HUMAN STANDARD: PRT: 1935 AA.  
AC P12883; Q14904; Q16579;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).  
GN MYH7 OR MYHCB.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91065634; PubMed-2249844;  
RA Jaencke T., Diederich K.W., Haas W., Schleich J., Lichter P.,  
RA Pfortl M., Bach A., Vosberg H.P.;  
RT "The complete sequence of the human beta-myosin heavy chain gene and  
a comparative analysis of its product.";  
RL Genomics 8:194-206(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-90301496; PubMed-2362820;  
RA Liew C.-C., Sole M.J., Yamauchi-Takahara K., Kellam B.,  
RA Anderson D.H., Lin L., Liew J.;  
RT "Complete sequence and organization of the human cardiac beta-myosin  
heavy chain gene.";  
RN [3]  
RN Nucleic Acids Res. 18:3647-3651(1990).

RP SEQUENCE OF 1-115 FROM N.A.  
RX MEDLINE-89264452; PubMed-2726733;  
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;  
RT "Characterization of human cardiac myosin heavy chain genes.";  
RN [4]  
RP ERRATUM.  
RA Yamauchi-Takahara K., Sole M.J., Liew J., Ing D., Liew C.-C.;  
RL Proc. Natl. Acad. Sci. U.S.A. 86:3504-3508(1989).  
RN [5]  
RP SEQUENCE OF 1310-1935 FROM N.A.  
RX MEDLINE-8616778; PubMed-2421254;  
RA Saez L., Leinwand L.A.;  
RT "Characterization of diverse forms of myosin heavy chain expressed in  
adult human skeletal muscle.";  
RN [6]  
RP Nucleic Acids Res. 14:2951-2969(1986).  
RN [7]  
RP REVISIONS.  
RA Leinwand L.A.;  
RL Submitted (MAR-1988) to the EMBL/Genbank/DBJ databases.  
RN [8]  
RP SEQUENCE OF 1410-1935 FROM N.A.  
RX MEDLINE-88299163; PubMed-2969919;  
RA Kurabayashi M., Tsuchimochi H., Komuro I., Takaku F., Yazaki Y.;  
RT "Molecular cloning and characterization of human cardiac alpha- and  
beta-form myosin heavy chain complementary DNA clones. Regulation of  
expression during development and pressure overload in human  
atrium.";  
RN [9]  
RL J. Clin. Invest. 82:524-531(1988).  
RN [10]  
RP SEQUENCE OF 785-1935 FROM N.A.  
RC TISSUE-Skeletal muscle;  
RX MEDLINE-90235862; PubMed-1691980;  
RA Bober E., Buchberger-Seitel A., Braun T., Singh S., Goeldie H.W.,  
RA Arnold H.H.;  
RT "Identification of three developmentally controlled isoforms of human  
myosin heavy chains.";  
RN [11]  
RL Eur. J. Biochem. 189:55-65(1990).  
RN [12]  
RP SEQUENCE OF 1393-1935 FROM N.A.  
RX MEDLINE-87192738; PubMed-3032769;  
RA Jandreski M.A., Liew C.-C.;  
RT "Construction of a human ventricular cDNA library and  
characterization of a beta myosin heavy chain cDNA clone.";  
RN [13]  
RL Hum. Genet. 76:47-53(1987).  
RN [14]  
RP REVIEW ON VARIANTS.  
RX MEDLINE-96039076; PubMed-8533830;  
RA Arai S., Matsuo K., Hirayama K., Sukurai H., Tamura M., Ozawa T.,  
RA Kimura M., Imamura S.-I., Furutani Y., Joh-O K., Kawana M., Takao A.,  
RA Hosoda S., Momma K.;  
RT "Missense mutation of the beta-cardiac myosin heavy-chain gene in  
hypertrophic cardiomyopathy.";  
RN [15]  
RL Am. J. Med. Genet. 58:267-276(1995).  
RN [16]  
RP VARIANTS CMH1 GLU-256 AND ARG-741.  
RX MEDLINE-93248216; PubMed-8483915;  
RA Fananapazir L., Dalakas M.C., Cyran F., Cohn G., Epstein N.D.;  
RT "Missense mutations in the beta-myosin heavy-chain gene cause central  
core disease in hypertrophic cardiomyopathy.";  
RN [17]  
RL Proc. Natl. Acad. Sci. U.S.A. 90:3993-3997(1993).  
RN [18]  
RP VARIANT CMH1 GLN-403.  
RX MEDLINE-9036731; PubMed-1975517;  
RA Geisterfer-Lowrance A.A.T., Kass S., Tanigawa G., Vosberg H.-P.,  
RA McKenna W., Seidman C.E., Seidman J.G.;  
RT "A molecular basis for familial hypertrophic cardiomyopathy: a beta  
cardiac myosin heavy chain gene missense mutation.";  
RN [19]  
RL Cell 62:999-1006(1990).  
RN [20]  
RP VARIANTS CMH1 GLN-249; GLN-403; CYS-453 AND MET-606.  
RX MEDLINE-92204193; PubMed-1552912;  
RA Watkins H., Rosenzweig A., Hwang D.S., Levi T., McKenna W.,

RA Seidmann C.E., Seidmann J.G.;  
 RT "Characteristics and prognostic implications of myosin missense  
 RT mutations in familial hypertrophic cardiomyopathy.";  
 RL New Engl. J. Med. 326:1108-1114(1992).  
 RN [14]  
 RP VARIANTS CMH1 GLN-403; CYS-453; ARG-584 AND MET-606.  
 RX MEDLINE-94070863; PubMed-8250038;  
 RA Watkins H., Thierfelder L., Anan R., Jarcho J., Matsumori A.,  
 RA McKenna W., Seidman J.G., Seidman C.E.;  
 RT "Independent origin of identical beta cardiac myosin heavy-chain  
 RT mutations in hypertrophic cardiomyopathy.";  
 RL Am. J. Hum. Genet. 53:1180-1185(1993).  
 RN [15]  
 RP VARIANTS CMH1 GLN-403 AND VAL-908.  
 RX MEDLINE-92346810; PubMed-1638703;  
 RA Epstein N.D., Cohn G.M., Cyran F., Fananapazir L.;  
 RT "Differences in clinical expression of hypertrophic cardiomyopathy  
 RT associated with two distinct mutations in the beta-myosin heavy chain  
 RT gene. A 908Leu->Val mutation and a 403Arg->Gln mutation.";  
 RL Circulation 86:345-352(1992).  
 RN [16]  
 RP VARIANTS CMH1 LEU-403 AND TRP-403.  
 RX MEDLINE-94075629; PubMed-8254035;  
 RA Dausse E., Komajda M., Fellet L., Dubourg O., Dufour C., Carrier L.,  
 RA Wisniewsky C., Bercovici J., Hengstenberg C., Al-Mahdawi S.;  
 RT "Familial hypertrophic cardiomyopathy. Microsatellite haplotyping and  
 RT identification of a hot spot for mutations in the beta-myosin heavy  
 RT chain gene.";  
 RL J. Clin. Invest. 92:2807-2813(1993).  
 RN [17]  
 RP VARIANTS CMH1 TRP-403.  
 RX MEDLINE-94093568; PubMed-8268932;  
 RA Moollan J.C., Brink P.A., Corfield V.A.;  
 RT "Identification of a new missense mutation at Arg403, a Cpg mutation  
 RT hotspot, in exon 13 of the beta-myosin heavy chain gene in  
 RT hypertrophic cardiomyopathy.";  
 RL Hum. Mol. Genet. 2:1731-1732(1993).  
 RN [18]  
 RP VARIANTS CMH1 ASN-615.  
 RX MEDLINE-93038688; PubMed-1417858;  
 RA Nishi H., Kimura A., Harada H., Toshima H., Sasazuki T.;  
 RT "Novel missense mutation in cardiac beta myosin heavy chain gene  
 RT found in a Japanese patient with hypertrophic cardiomyopathy.";  
 RL Biochem. Biophys. Res. Commun. 188:379-387(1992).  
 RN [19]  
 RP VARIANTS CMH1 GLY-778.  
 RX MEDLINE-93343938; PubMed-8343162;  
 RA Harada H., Kimura A., Nishi H., Sasazuki T., Toshima H.;  
 RT "A missense mutation of cardiac beta-myosin heavy chain gene linked  
 RT to familial hypertrophic cardiomyopathy in affected Japanese  
 RT families.";  
 RL Biochem. Biophys. Res. Commun. 194:791-798(1993).  
 RN [20]  
 RP VARIANTS CMH1 VAL-908.  
 RX MEDLINE-93168485; PubMed-8435239;  
 RA Al-Mahdawi S., Chamberlain S., Cleland J., Mihoyannopoulos P.,  
 RA Gilligan D., French J., Choudhury L., Williamson R., Oakley C.;  
 RT "Identification of a mutation in the beta cardiac myosin heavy chain  
 RT gene in a family with hypertrophic cardiomyopathy.";  
 RL Br. Heart J. 69:136-141(1993).  
 RN [21]  
 RP VARIANTS CMH1 TRP-719.  
 RX MEDLINE-95179132; PubMed-7874131;  
 RA Greve G., Bachinski L., Friedman D.L., Czeruzewicz G., Anan R.,  
 RA Towbin J., Seidman C.E., Roberts R.;  
 RT "Isolation of a de novo mutant myocardial beta MHC protein in a  
 RT pedigree with hypertrophic cardiomyopathy.";  
 RL Hum. Mol. Genet. 3:2073-2075(1994).  
 RN [22]  
 RP VARIANTS CMH1 CYS-513; ARG-716 AND TRP-719.  
 RX MEDLINE-94110336; PubMed-8282798;  
 RA Anan R., Greve G., Thierfelder L., Watkins H., McKenna W., Solomon S.,  
 RA Vecchio C., Shono H., Nakao S., Tanaka H., Mares A. Jr., Towbin J.A.,

RA Spirito P., Roberts R., Seidman J.G., Seidman C.E.;  
 RT "Prognostic implications of novel beta cardiac myosin heavy chain gene  
 RT mutations that cause familial hypertrophic cardiomyopathy.";  
 RL J. Clin. Invest. 93:280-285(1994).  
 RN [23]  
 RP VARIANTS CMH1 THR-797.  
 RX MEDLINE-96047159; PubMed-7581410;  
 RA Moollan J.C., Brink P.A., Corfield V.A.;  
 RT "Identification of a novel Ala797Thr mutation in exon 21 of the beta-  
 RT myosin heavy chain gene in hypertrophic cardiomyopathy.";  
 RL Hum. Mutat. 6:197-198(1995).  
 RN [24]  
 RP VARIANTS CMH1 CYS-453.  
 RX MEDLINE-96209901; PubMed-8655135;  
 RA Ko Y.-L., Chen J.-J., Tang T.-K., Cheng J.-J., Lin S.-Y., Liou Y.-C.,  
 RA Kuan P., Wu C.-W., Lien W.-P., Liew C.-C.;  
 RT "Malignant familial hypertrophic cardiomyopathy in a family with a  
 RT 453Arg->Cys mutation in the beta-myosin heavy chain gene:  
 RT coexistence of sudden death and end-stage heart failure.";  
 RL Hum. Genet. 97:585-590(1996).  
 RN [25]  
 RP VARIANTS CMH1 THR-349 AND TRP-719.  
 RX MEDLINE-98204402; PubMed-9544842;  
 RA Jeschke B., Uhl K., Weist B., Schroder D., Metlinger T.,  
 RA Dohlemann C., Vosberg H.-P.;  
 RT "A high risk phenotype of hypertrophic cardiomyopathy associated with  
 RT Query Match 5.2%; Score 123; DB 1; Length 1935;  
 RT Best Local Similarity 22.1%; Pred. No. 6;  
 RT Matches 94; Conservative 63; Mismatches 140; Indels 128; Gaps 19;  
 QY 7 QIEQLSKPL-----SDSICGYLKK--LEKSAFRLRNEF-----NVAQTALR 47  
 DB 1385 ELERAKKKLAQRLOEAEVAEVAANKCSLEKTKHR-LQNEIEDLVVERSNAAAALD 1443  
 QY 48 KLSQNK-----PSADERDALQKAC-----LKKKRLTSD 74  
 DB 1444 KQKNFDFKILAEWKQKYEEQSQSELESSQKARSSTELFKLNAYEESLEHLETFKRENK 1503  
 QY 75 SLVEQFS-----KTRDIELISFVAAQFL-LDTTLESANSLFWLADLSEKHMWD 123  
 DB 1504 NLQPEISDLIEBOLGSSKKTITHELEKVKQKLEAEKMELOSALKEAEASLE-----HEEGK 1557  
 QY 124 HLNVPVLEPVELTKSD-----DDKGEREQA--DAKYAFQVLGVDSSESSILYAPVLQLP 175  
 DB 1558 ILRQLEFQNIKAIERKLEKDEBQAKRNHLRVVDSIQTSIDATRSRNEALRYKKR 1617  
 QY 176 LVGEVTFEFDQ-----SAERKEISQLKSMLTITVAQ-----ERRAT----- 212  
 DB 1618 MEGDLNEMETQLSHANRMAEAQKQVKSLOSLSLKDQIOLDDAVRANDLKEMIAIERR 1677  
 QY 213 ----QFKMENAKRCVTOLDRLSAL-----VSTKCHSLGASQSNFGFAK-----SL 253  
 DB 1678 NNLLQAELEELRAVEDETERSKRLAEQELIETSERVOLLQSNTSLNQKKKDDADLSOL 1737  
 QY 254 LTRVENALVHLGKILKAPKAQKTEQVEAESSVSEGELEPSPMDTK-HIERIPMAEQAO 312  
 DB 1738 QTEVEEA-----VQECRNMEK-AKKALITDAAMMADELKKEQDTSNHLERMKKNMQTI 1790  
 QY 313 TVSQH 317  
 DB 1791 KDLQH 1795  
 RESULT 9  
 MYH7\_RAT  
 ID MYH7\_RAT STANDARD; PRT; 1935 AA.  
 AC P02564;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Myosin heavy chain, cardiac muscle beta isoform (MYHC-beta).  
 GN MYH7.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RX MEDLINE=90016823; PubMed=2798112;  
 RA Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;  
 RT "Complete nucleotide sequence of full length cDNA for rat beta  
 cardiac myosin heavy chain."  
 RL Nucleic Acids Res. 17:7529-7530(1989).  
 RN  
 RP DISCUSSION OF SEQUENCE.  
 RX MEDLINE=90133919; PubMed=2614840;  
 RA McNally E.M., Kraft R., Bravo-Zehnder M., Taylor D., Leinwand L.A.;  
 RT "Full-length rat alpha and beta cardiac myosin heavy chain sequences.  
 Comparisons suggest a molecular basis for functional differences."  
 J. Mol. Biol. 210:665-671(1989).  
 RN  
 RP SEQUENCE OF 1524-1935 FROM N.A.  
 RX MEDLINE=82220036; PubMed=7045682;  
 RA Mahdavi V., Periasamy M., Nadal-Ginard B.;  
 RT "Molecular characterization of two myosin heavy chain genes expressed  
 in the adult heart."  
 RL Nature 297:659-664(1982).  
 RN  
 RP SEQUENCE OF 1871-1935 FROM N.A.  
 RC STRAIN=Histar; TISSUE=Heart;  
 RX MEDLINE=85179510; PubMed=6241892;  
 RA Mahdavi V., Lompre A.M., Chambers A.P., Nadal-Ginard B.;  
 RT "Cardiac myosin heavy chain isozymic transitions during development  
 and under pathological conditions are regulated at the level of mRNA  
 availability."  
 RL Eur. Heart J. 5:181-191(1984).  
 CC  
 CC -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
 CC -1- P.TM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY  
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.  
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE  
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
 SUBFRAGMENT (S2).  
 CC -1- MISCELLANEOUS: THE CARDIAC ALPHA ISOFORM IS A 'FAST' ATPASE  
 MYOSIN, WHILE THE BETA ISOFORM IS A 'SLOW' ATPASE.  
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: X15939; CAA34065.1; -  
 CC EMBL: J00752; AAA41654.1; -  
 CC EMBL: M32698; AAA41659.1; -  
 CC PIR: S06006; S06006.  
 CC PIR: A02989; A02989.  
 CC HSSP: P08799; IMD.  
 CC InterPro: IPR000048; IQ\_region.  
 CC InterPro: IPR004009; Myosin\_N.  
 CC InterPro: IPR002928; Myosin\_tail.  
 CC InterPro: IPR001609; myosin\_head.  
 CC Pfam: PF00063; myosin\_head; 1.

DR Pfam: PF00612; IQ; 2.  
 DR Pfam: PF01576; Myosin\_tail; 1.  
 DR Pfam: PF02736; Myosin\_N; 1.  
 DR PRINTS: PR00193; MYOSINHEAVY.  
 DR PRODOM: PD000355; myosin\_head; 1.  
 DR SMART: SM00015; IQ; 1.  
 DR SMART: SM00242; MYSC; 1.  
 DR PROSITE: PS50096; IQ; 1.  
 KW Myosin; muscle protein; coiled coil; Thick filament; Actin-binding;  
 KW ATP-binding; Methylation; Alkylation; Multigene family;  
 KW Calmodulin-binding.  
 FT DOMAIN 781 800 MYOSIN HEAD-LIKE.  
 FT DOMAIN 810 810 IQ.  
 FT DOMAIN 840 1935 COILED COIL (POTENTIAL).  
 FT NP\_BIND 178 185 ATP.  
 FT DOMAIN 655 677 ACTIN-BINDING.  
 FT DOMAIN 757 771 ACTIN-BINDING.  
 FT MOD\_RES 129 129 METHYLATION (TR1-) (POTENTIAL).  
 FT MOD\_RES 695 695 ALKYLATION (SH-1) (POTENTIAL).  
 FT MOD\_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).  
 FT CONFLICT 1529 1531 TRK -> VRR (IN REF. 3).  
 FT CONFLICT 1731 1731 D -> H (IN REF. 3).  
 FT CONFLICT 1784 1784 N -> K (IN REF. 3).  
 FT CONFLICT 1851 1851 T -> N (IN REF. 3).  
 FT CONFLICT 1858 1858 R -> K (IN REF. 3).  
 SQ SEQUENCE 1935 AA; 223082 MW; C8376C324A7BD82B CRC64;

Query Match 5.2%; Score 123; DB 1; Length 1935;  
 Best local Similarity 22.1%; Pred. No. 6;

Matches 102; Conservative 65; Mismatches 154; Indels 140; Gaps 21;

OY 7 QIQOLSKPL-----SDDSICGYLLK--LEKSAFPLRNEF-----NVAOTLAIR 47  
 DB 1385 ELDEAKKKLAQRLQDADEAEVAVNAKSSLEKTKHR-LQNEIEDLWVYDERSAAAAAALD 1443  
 OY 48 KLSQN-----PSADERDALQENAC-----LNKMKILSD 74  
 DB 1444 KQGNDFKILVEMKQYEESSQSESSQKARSLSLEFLKNAVEESLEHLETFPRENK 1503  
 OY 75 SLVEQFS-----KTRDIELISWYVAQFL-LDTTLESAANSLEWLADLSKHM 123  
 DB 1504 NLOEELSDLEQSGTSGSIHELEKLRKOLEAKLELSQALEEVAESL-----HEEGK 1557  
 OY 124 HLNPVLPVETLKSD-----DDKGEREQA---DAKYAFPOLGDSSESSILYAPVLOLP 175  
 DB 1558 ILQAOLEFNQIKAEIERKLAKEDEMEQAKRNHLRVYDSIQTSLDQFTRSNALAVKKK 1617  
 OY 176 LVGEVTFEFDQ-----SAERKEISQLKSMLTJTVAQ-----ERFAI---- 212  
 DB 1618 MEGDLNMEETQLSHANMAAEQAQVKSLSQSLKDPQIQOLDVAVRANDLKENIAIVERR 1677  
 OY 213 -----QKMEAKKCVQLDRLSL-----VTKCHSLGSGSTNGCFAP-----SL 253  
 DB 1678 NNLLQAELEBELRAVVBQTESRKLAQELITSEERQOLLSQNTSLNOKKKMDADLSQI 1737  
 OY 254 LTFVENALVLSGIRKLAPKAEKTVQEOVAESSVSGELPSHDTK-HIERIPMASEOAO 312  
 DB 1738 QTEVEEA-----VQECRNAEEK-AKKAITDAAMAEELKKEDDTJSAHLER--MKNNMED 1788  
 OY 313 TVSQHLL-----AGNLSELGNLNMNNDLAFLHLE 343  
 DB 1789 TIKDLQRIQDAEQIALKGKQKQLOKLEARVRELEMLEAE 1829

RESULT 10  
 ANTI\_ONCVO  
 ID ANTI\_ONCVO STANDARD; PRT: 2022 AA.  
 AC P21249;  
 DT 01-MAY-1991 (rel. 18, Created)  
 DT 01-NOV-1995 (rel. 32, Last sequence update)  
 DT 30-MAY-2000 (rel. 39, Last annotation update)  
 DE Major antigen.  
 GN OVT1.



OS Onchocerca volvulus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
OC Onchocercidae; Onchocerca.  
OX NCBI\_TaxId=6282;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95287898; PubMed=7770081;  
RA Trilteerpregh S., Richle T.L., Tuan R.S., Shepley K.J., Dinman J.D.,  
RT Neuber T.A., Scott A.L.;  
RT "Molecular cloning of a gene expressed during early embryonic  
RT development in Onchocerca volvulus."  
RL Mol. Biochem. Parasitol. 69:161-171(1995).  
RN  
RP SEQUENCE OF 733-866 FROM N.A.  
RX MEDLINE=89127417; PubMed=2464764;  
RA Donaldson J.E., Duke B.O.L., Moser D., Zeng W., Erondou N.E.,  
RT Lucius R., Renz A., Karam M., Flores G.Z.;  
RT "Construction of Onchocerca volvulus cDNA libraries and partial  
RT characterization of the cDNA for a major antigen."  
RL Mol. Biochem. Parasitol. 31:241-250(1988).  
CC -1- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U12681; AAA80009.1; -;  
DR EMBL: J03995; AAA29412.1; -;  
DR PIR: A54513; A54513.  
KW Antigen; Coiled coil.  
FT DOMAIN 74 120 COILED COIL (POTENTIAL).  
FT DOMAIN 151 251 COILED COIL (POTENTIAL).  
FT DOMAIN 327 384 COILED COIL (POTENTIAL).  
FT DOMAIN 417 1879 COILED COIL (POTENTIAL).  
SQ SEQUENCE 2022 AA; 237341 MW; B7132AACF1520317 CRC64;  
  
Query Match 5.2%; Score 123; DB 1; Length 2022;  
Best Local Similarity 22.9%; Pred. No. 6.4; Indels 82; Gaps 16;  
Matches 85; Conservative 64; Mismatches 140;  
  
QY 35 LRNEFVAQTALRKLSQNPASDERDALOEACLNKWKILSDSLYEQFSKTRDIELISFEV 94  
DB 971 LRKSLNDAERAMADL-QN-----RDSILERENDMKREKSDALNNEIDLR-LRD-ELLSVRR 1022  
QY 95 AAOFL- -DTTESAA-NSLEWLADISEKHMDHNLNVLFEVETLKSDDDKGKERQADAK 150  
DB 1023 DAEKEINRYTDTQTAARNEIKLLPTNNEMKSQLN-----AAEDK 1063  
QY 151 VKAFPOLVGSESSSILYAVLQPLVGEYTFDFQSAERKGEISOLKSMLTVTVAERF 210  
DB 1064 INSLNKVITQONKI-----RDLTGEVHHLEGELKDKGNVANLESELDPT- -RE 1111  
QY 211 AIOFKMENARCYTQDRLSALVSTKCHSLGOSTNFGFAKSLTTRVENALVHLSGIKLA 270  
DB 1112 RIHLGQNSLSQTELKIK-----GDIDSLFG-----ENM- LTKAKES 1150  
QY 271 PKAEAKTVEEVAESSSEELPSHMDTKHIERIPMASEOQVYSHLHAGNLSLGNLN 330  
DB 1151 NEAEIRLKKOKLORSTENAKKYSADLKLPEYDRL-----QMLYREKIKQAENLT 1201  
QY 331 NMRDLAFHL- -LREYSDYFROSE- PHSPIFLLEKAIKMGYSLBELLEKRMNSEQ 383  
DB 1202 QAVVDLESRLNOSRRRLRDTKTLASEGDRNALRSEVERLQHEVQPMRQLLKK- -TDE 1259  
QY 384 NGDALSTIFNA 394  
DB 1260 YQAALSLVANA 1270

RESULT 11  
TEGU\_HSVSA STANDARD; PRT; 2469 AA.  
ID TEGU\_HSVSA  
AC 001056;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DE 01-APR-1993 (Rel. 25, Last annotation update)  
DE Probable large tegument protein.  
GN 64 OR ERF2.  
OS Herpesvirus saimiri (strain 11).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Camaherpesvirinae; Rhadinovirus.  
OX NCBI\_TaxId=10383;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92333688; PubMed=1321287;  
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,  
RA Honess R.W.;  
RT "Primary structure of the herpesvirus saimiri genome."  
RL J. Virol. 66:5047-5058(1992).  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92230228; PubMed=1314457;  
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;  
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of  
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic  
RT organization between HVS and Epstein-Barr virus."  
RL Virology 188:286-310(1992).  
CC -1- FUNCTION: TEGUMENT PROTEIN.  
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,  
CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X64346; CAA55687.1; -;  
DR EMBL: M66409; AAA46140.1; -;  
DR PIR: H36812; H36812.  
SQ SEQUENCE 2469 AA; 280165 MW; D2B4B8DC08644CDB CRC64;  
  
Query Match 5.2%; Score 122.5; DB 1; Length 2469;  
Best Local Similarity 20.5%; Pred. No. 9; Indels 221; Gaps 27;  
Matches 121; Conservative 77; Mismatches 172;  
  
QY 40 NVAQOTALRKLSQ-----NPSADERDALO-----EACLNKWKILSDSLYEQFSKTRR 85  
DB 694 NVLTDLHTKKOKTOISPAFDFTYVKNIGSTLQGLQTCVTDLNT-DKFF- -ISNTVQ 749  
QY 86 DIELISFVAQFLDLTLESAAANSLEWLADISEKHMD- -HLNPLVFEVETLKSDDDKGE 143  
DB 750 QLSYIGWEV-----AELSHSQNPFKADVPVLPKLIT- -DDIKKE 786  
QY 144 RQADAKVK-----AFQO-LVGDSESSSILYAVLQPL- -PLVGEYTFDFQES 187  
DB 787 IQOVTTKQKEEPLSKTLAVQVOTLENAKOSDTLSIPDIQHYTTKAGTLGGERENOKFES 846  
QY 188 AE- -RKGEISQ- -LKSMLTVAQERFAIOFMENAKKRCVTDRLSALVSTKCHSLGS 242  
DB 847 LKNTVQKLSSTSEELKTLIDST-----LENVQ- -LQIGESIDLSQNOYIHQS 993  
QY 243 QSTNFGFAKSLTTRVENALVHLSGIK-----LAPK- -AEAKTVEO- -EVAESS 286  
DB 894 ETIKQAFEDKSNITINNIIQILNQOKYTTVTOPMLIAVKRFLSEAKFRESENTICEIISTL 953  
QY 287 VSEGELPSHMDTKHIERIPMASRQAOVSHLHAGNLSLGNLNMRDLAFHL- -LREY 344  
DB 954 VSLGSLSKSTT- -VEALKDALKSIDTLKERTLAVDPRPKRELRYNIRKLOKOLKPTLLQ 1011



QY 345 SDY-----FROSEPHSPISFLEKAIKRWGLSL-----PELLREM 379  
Db 1012 QEDDNKMEVDSPVPTPSRDVKTFFIONAPSMKAKQYAKKALKOIOAMEDVDPESVIED 1071  
QY 380 MSQONGDAL-----STIFNAG-----386  
Db 1072 NIKANGOKAMQKIQSAFODLNFSILIPDDWLSLAKEYTRKPSLTFTVIGPILKFEVEVL 1131  
QY 397 -----LN-HLDVULPREVST--PRVGIESP 418  
Db 1132 ESKVNLKEAKLKLLPNGPVFTPEPKEDWIHYESNVNHLKTLNLPKSVTVAHNGHEL 1191  
QY 419 QTFQAKPSVDP-----SVEEHVSQTS-----PDTOSKOD 451  
Db 1192 LLSQALNSKLTPEAVYGTSLDQHAAKFSKFKLTLEATWHDHVDTRKIDE 1242

RESULT 12  
MYH7\_MESAU  
ID MYH7\_MESAU STANDARD: PRT: 1934 AA.  
AC P13540: 060540:  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE MYOSIN heavy chain, cardiac muscle beta isoform (MyHC-beta).  
GN MYH7.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OX Mesocricetus.  
NCBI\_TaxID=10036;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=F1B; TISSUE=Liver;  
RX MEDLINE=95115033; PubMed=7815459;  
RA Wang R., Sole M.J., Cukerman E., Liew C.-C.;  
RT "Characterization and nucleotide sequence of the cardiac alpha-myosin  
heavy chain gene from Syrian hamster";  
RL J. Mol. Cell. Cardiol. 26:1155-1165(1994).  
RN (2)  
RP SEQUENCE OF 962-1935 FROM N.A.  
RX MEDLINE=88247788; PubMed=3380703;  
RA Jandreski M.A., Sole M.J., Liew C.-C.;  
RT "Sequence of cDNA encoding the Syrian hamster cardiac beta-myosin  
heavy chain";  
RL Nucleic Acids Res. 16:4737-4737(1988).  
CC -1- FUNCTION: MUSCLE CONTRACTION.  
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2  
HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)  
AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.  
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING  
CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,  
CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.  
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT  
MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE  
SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED  
SUBFRAGMENT (S2).  
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation-  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L12104; AA02313.1; -  
CC EMBL: X07273; CAA30256.1; -  
CC PIR: A28298; A28298.

DR HSP; P08799; 1MDN.  
DR InterPro: IPR000048; IQ\_region.  
DR InterPro: IPR004009; Myosin\_N.  
DR InterPro: IPR002928; Myosin\_tail.  
DR InterPro: IPR001609; myosin\_head.  
DR Pfam: PF00633; myosin\_head.  
DR Pfam: PF00612; IQ\_2.  
DR Pfam: PF01576; Myosin\_tail.  
DR Pfam: PF02736; Myosin\_N.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR ProDom: PD000355; myosin\_head.  
DR SMART: SM00015; IQ\_1.  
DR SMART: SM00242; MYSC\_1.  
DR PROSITE: PS50096; IQ\_1.  
KW Myosin; Muscle protein; coiled coil; Thick filament; Actin-binding;  
KW ATP-binding; Methylation; Alkylation; Multigene family;  
KW Calmodulin-binding.  
FT DOMAIN 1 779 MYOSIN HEAD-LIKE.  
FT DOMAIN 780 809 IQ.  
FT DOMAIN 839 1934 COILED COIL (POTENTIAL).  
FT NP\_BIND 177 184 APP.  
FT DOMAIN 654 676 ACTIN-BINDING.  
FT DOMAIN 756 770 ACTIN-BINDING.  
FT MOD\_RES 128 128 METHYLATION (TR1-) (POTENTIAL).  
FT MOD\_RES 694 694 ALKYLATION (SH-1) (POTENTIAL).  
FT MOD\_RES 704 704 ALKYLATION (SH-2) (POTENTIAL).  
FT CONFLICT 966 966 D -> E (IN REF. 2).  
FT CONFLICT 978 978 T -> TE (IN REF. 2).  
FT CONFLICT 986 986 E -> Q (IN REF. 2).  
FT CONFLICT 1008 1014 DLQAEED -> ALEARKT (IN REF. 2).  
FT CONFLICT 1057 1057 D -> Y (IN REF. 2).  
FT CONFLICT 1060 1060 L -> V (IN REF. 2).  
FT CONFLICT 1095 1095 D -> N (IN REF. 2).  
FT CONFLICT 1217 1217 E -> D (IN REF. 2).  
FT CONFLICT 1271 1271 D -> N (IN REF. 2).  
FT CONFLICT 1327 1327 T -> A (IN REF. 2).  
FT CONFLICT 1358 1358 C -> R (IN REF. 2).  
FT CONFLICT 1504 1504 L -> V (IN REF. 2).  
FT CONFLICT 1537 1537 M -> L (IN REF. 2).  
FT CONFLICT 1556 1556 N -> K (IN REF. 2).  
SQ SEQUENCE 1934 AA; 222928 MW; FDBAC58310B0B57D CRC64;

Query Match 5.2%; Score 122; DB 1; Length 1934;  
Best Local Similarity 2.1%; Pred. No. 77;  
Matches 94; Conservative 64; Mismatches 139; Indels 128; Gaps 19;

QY 7 QIEQLSKPL-----SDSICGVYLK--LEKSAFRLPRLNEF-----NYAQTLR 47  
Db 1384 ELEBAKKKLAQRLQDAEAEVAVNAKSSLEKTKRHLQNEIEDLVYDERSNMAAALD 1442  
QY 48 KLSQN-----PSADERDALQEC-----LNKWLSD 74  
Db 1443 KRONPDKILAEWKQKYEESSQSESSQKPARSLSTELFKNAVEESLEHLETFKRENK 1502  
QY 75 SLYEORS-----KTRDIELISWFAVLAQFL-LDTLLESANSLWLAJLSEKHW 123  
Db 1503 NIDBEISDLTQGLSGTGSSTHELEKIRKOLEAEKMEQLSELEBESLSE-----HEGN 1556  
QY 124 HLNPLVPLETFLKSD-----DQKGEREA--DAKVAAPFOLGVDSSESSILYAPVLDLP 175  
Db 1557 ILRQLEFNQIKAEIERKLEKDEMEQAKRNHLRVVDSLQTSLDATRSRNEALRYKKK 1616  
QY 176 LVGEVYFFPDQ-----SARKGEISQLKSMLTYYAQ-----ERFAT-----212  
Db 1617 MEDDLKMEIQLSHANMAEAQKQVKSLOSLLKTDQIQDDAVRANDLKEMAIYERR 1676  
QY 213 -----OFKEMAKPCVQLDLSAL-----VSTKCHSLGSSSTNFGFAK-----SL 253  
Db 1677 NNLQAELEELRAVNVDETERSKRLADELLETSEKROLHSNTSLINOKKKKADADLSOL 1736  
QY 254 LTRVENALVHLSCIKLAPKRAKTAQTVQEVAVSSVSGELPSSHDTK-HIERIPMASBOAQ 312  
Db 1737 QTEVEEA-----VQECRNAAEK-AKKAITDAMMAEELKKEDDTSAHLERMKKNBDTI 1789

QY 313 TVSOH 317  
 DB 1790 KDLOH 1794

RESULT 13  
 OXRP\_HUMAN  
 ID OXRP\_HUMAN STANDARD; PRT; 999 AA.  
 AC OXVALI:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 150 kDa oxygen-regulated protein precursor (Orp150) (Hypoxia up-regulated 1).  
 GN HYOU1 OR Orp150.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Astrocytoma;  
 RX MEDLINE=97148379; PubMed=9020069;  
 RA Ikeda J., Kaneda S., Kuwabara K., Ogawa S., Kobayashi T., Matsumoto M., Yura T., Yanagi H.;  
 RT "Cloning and expression of cDNA encoding the human 150 kDa oxygen-regulated protein, Orp150.";  
 RL Biochem. Biophys. Res. Commun. 230:94-99(1997).  
 RN [2]  
 RP FUNCTION.  
 RX MEDLINE=99156921; PubMed=10037731;  
 RA Ozawa K., Kuwabara K., Tamatani M., Takatsuki K., Tsukamoto Y., Kaneda S., Yanagi H., Stern D.M., Eguchi Y., Tsujimoto Y., Ogawa S., Tohyama M.;  
 RT "150 kDa oxygen-regulated protein (Orp150) suppresses hypoxia-induced apoptotic cell death.";  
 RL J. Biol. Chem. 274:6397-6404(1999).  
 CC -1- FUNCTION: HAS A PIVOTAL ROLE IN CYTOPROTECTIVE CELLULAR MECHANISMS TRIGGERED BY OXYGEN DEPRIVATION. MAY PLAY A ROLE AS A MOLECULAR CHAPERONE AND PARTICIPATE IN PROTEIN FOLDING.  
 CC -1- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.  
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TISSUES THAT CONTAIN WELL-DEVELOPED ENDOPLASMIC RETICULUM AND SYNTHESIZE LARGE AMOUNTS OF SECRETORY PROTEINS. HIGHLY EXPRESSED IN LIVER AND PANCREAS AND LOWER EXPRESSION IN BRAIN AND KIDNEY. ALSO EXPRESSED IN MACROPHAGES WITHIN AORTIC ATHEROSCLEROTIC PLAQUES, AND IN BREAST CANCERS.  
 CC -1- INDUCTION: BY HYPOXIA AND ALSO BY 2-DEOXYGLUCOSE OR TUNICAMYCIN.  
 CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U65785; AAC50947.1; -  
 DR Genew; HGNC:16931; HYOU1.  
 DR MIM: 601746; -  
 DR InterPro; IPR001023; Hsp70.  
 DR Pfam: PF00012; Hsp70; 2.  
 DR PRINTS: PR00301; HEATSHOCK70.  
 DR Prodom: PD000089; Hsp70; 2.  
 DR PROSITE; PS00297; HSP70\_1; PALSE\_NEG.  
 DR PROSITE; PS00329; HSP70\_2; 1.  
 DR PROSITE; PS01036; HSP70\_3; 1.  
 KW APP-binding; Chaperone; Endoplasmic reticulum; Signal.  
 FT SIGNAL 1 32  
 FT CHAIN 33 999  
 FT DOMAIN 603 606  
 POLY-GLU.

FT DOMAIN 636 641 POLY-PRO.  
 FT CARBOHYD 155 155 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 596 596 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 830 830 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 862 862 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 869 869 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 922 922 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 931 931 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 996 999 PREVENT SECRETION FROM ER (POTENTIAL).  
 SQ SEQUENCE 999 AA; 111335 MW; FCE0F292466AFAB9 CRC64;

Query Match 5.1%; Score 120.5; DB 1; Length 999;  
 Best Local Similarity 22.2%; Pred. No. 3.6;  
 Matches 89; Conservative 63; Mismatches 122; Indels 127; Gaps 19;

QY 81 SKTTTRDIELISWFAAOFLLDTTLESANSLIEMLADLSEKHWHLNPVLETLSDDK 140  
 DB ARKRMEVEIGVELVLDLPDLPDKLAQSVOKRQDL-----TLR--DLE 732

QY 141 GKREKADAKYKAFQVLGDSSESIYAPVQLPLVGEVTFPQSAERKGETLSQSKM 200  
 DB 733 KQREKANSLEAFI-----FETDKLYQPEYQ-----EVS-TEQREISGKLSASTW 781

QY 201 L-----TTTVAQERFA-----IQFMENAKRCVTQDLRLSALVSTKCHSLGQSQT 245  
 DB 782 LDEGCGATTYMLKEKLAELRLKLOGLEFPEREKKRP-----ERLSALDNLNHS----- 832

QY 246 NEGFAKSLTRVENALVHLSGIKLAPKAEA--KTVEQVAAESSVSEGLPSHMDTKIIE 303  
 DB 833 -----SMFLKGARLIPENDQIFTEVEVMTTEKVENIETWAMKNATLADQAK 877

QY 304 IPMSAQQTYSOHLHNGNLSELGNLNNMNDLAFHLRLREVSDFRQSEPHSPISFLFK 363  
 DB 878 LP-ATEKPVLLSKDIEAKMA-----LDREV-----QYLNK 908

QY 364 AIRNGYSLPELLREMMSEQNDALSTIFNAAGLNHLDOVLPVSPPTVGIESPQPOA 423  
 DB 909 A-----KTKTP---RPRKDKNGTRAEPPLNMSASDQGEKVP-----AGTEDA 951

QY 424 KPSVSDPSVEHVSQTSVPDQ-----SKDQKPOSS 456  
 DB 952 EP-ISEPEKVE---TGSBPQDTEPLELGGPAEPEQEQST 988

RESULT 14  
 MYH4\_RABIT  
 ID MYH4\_RABIT STANDARD; PRT; 1938 AA.  
 AC Q28641;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Myosin heavy chain, skeletal muscle, juvenile.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;  
 RA Maeda K., Hostilnova E., Roesch-Kleinkauf A., Schuster H., Gasperik J., Wittinghofer A.;  
 RT "Isolation, sequencing of myosin heavy chain cDNA from rabbit skeletal muscle and a novel cosynthesis of S-1 fragment with the essential and regulatory light chains.";  
 RT Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
 RL -1- FUNCTION: MUSCLE CONTRACTION.  
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).  
 CC -1- SUBCELLULAR LOCATION: THICK FILAMENTS OF THE MYOFIBRILS.  
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

```

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEMOYOSIN (LM) AND 1 HEAVY MEMOYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (SI) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U32574; AAA74199.1; -.
DR HSSP: P13538; 2MYS.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR000193; MYOSINHEAVY.
DR PRODOM: PD00035; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
DR KMW: Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 783 MYOSIN HEAD-LIKE.
FT DOMAIN 1 784 813 IQ.
FT DOMAIN 1 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 698 698 ALKYLATION (SH-1) (BY SIMILARITY).
FT MOD_RES 708 708 ALKYLATION (SH-2) (BY SIMILARITY).
SQ SEQUENCE 1938 AA; 223064 MW; D8A8A2EC5B182626 CRC64;

Query Match
Best Local Similarity 21.3%; Score 120.5; DB 1; Length 1938;
Matches 112; Conservative 87; Mismatches 196; Indels 131; Gaps 25;

OY 7 QIEQLSKPLSD-----DSICGYLLKLEKSAFRLRNEF-----NVAQTALR 47
DB 1388 ELEEAKKLIARLQDAEEHVAVNACASLEKTKQR-LQNEVEDLMIDVERTAAACALID 1446
OY 48 KLSQNSADERADLQACLNKKKILSDLYEOPSKTTRIOELISFVAAOFLLDTLLESA 107
DB 1447 KQQRN-----FDKILAEKHKKEETHAELASQKRSRSLSTEV---FKVKNAYEES 1494
OY 108 ANSLLEWL-----ADLSKHWHDHLPVLPVETLKSDDKGER-----EODARY 151
DB 1495 LQOLETLKRNKNNLQOEISDLTEQINAGSKRIHELKVKQYQOESELOALEEKEASI 1554
OY 152 KAFPOLVGSSESSILYAPVLOPLVGEVTFDFQSAEKKEISQLK-----SMLTT 203
DB 1555 -----EHEGKILR---IOLELNQVSEIDRIKIAEKDEIDQKRNHIRVESMOQT 1603
OY 204 TVAGER-----FAIQFKME-NAKRCVTOUDRLSALVSTCHSIGSOSTNFGPKSLITRY 257
DB 1604 LDEIIRNDIAIRIKKMEGLDNEIEIOLNHNRMMA---EALIRNYRNTQIGLKDTQLHL 1660

```

```

OY 258 ENALVHLGKIK--LAD-----KAERK-----TVEQVAESSVSGELPSHMDTKHIER 303
DB 1661 DDLRQGEDLKEQLAVERRANLQAEIETELRATLQGTESKRVAQEL---LD----- 1711
OY 304 IPWASQAOQVSOHLHAGNLSELGNLNNMNRDLAPHLRLVSDYFPQSEPHSPISFLER 363
DB 1712 ---ASERVOL---LHTQNTSLINTFKKLETDTIS-OIOGEMEDIVQEARNAEKA---RK 1760
OY 364 AIRWGLSLPELLREMSQNGDALSTIF--NAAGLNH-LD-----QVLLP 406
DB 1761 ALTDAAMAEELKEEDDTSAHLERMKKNMEQTVKDIQHRLEDEAQLGKKRQIKQLLEA 1820
OY 407 EVSTPVGIESPOTPAKPSVDPKRSVEHVSQTSVPYDQSKODK 452
DB 1821 RVAELEAEVSESG---KKNVEAVKGLRKHRRVKELTYQTEDRK 1862

RESULT 15
M10.CAEEL
ID M10.CAEEL STANDARD: PRT; 650 AA.
AC P34400;
DT 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE M19-10 protein.
GN M19-10 OR F10E9.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten T., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders E., Shownkeen R.,
RA Sims M., Smallton N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thiermy-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Woldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L10986; AAA28018.1; -.
DR PIR: S44806; S44806.
DR Wormpep: F10E9.6; CE00150.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000159; RA_domain.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00314; RA; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
FT DOMAIN 327 437 PH.
SQ SEQUENCE 650 AA; 73318 MW; C3532BBF587D1335 CRC64;

Query Match
Best Local Similarity 19.5%; Score 120; DB 1; Length 650;
Matches 101; Conservative 85; Mismatches 208; Indels 124; Gaps 24;

```

```

QY 7 QIBQLSKPLSDSICGVYKL---EKSAPRPLRNE-FNVAQTALRKLSONPSADERDALQ 62
Db 178 KIROALEKMEKAVTKIFVFEVDEGALQOLLIDERVTADTLKQLAEKNHIA---LME 233
QY 63 EACLINRW--KILSDSLYEQSKTTTROIELISW-----FVAAQFLDIT 102
Db 234 DHCIVEEYPELLYTKRYEDHEKVENIOM--WQODSPNKLYEMRRDPDKYAFISRPELYLL 291
QY 103 TLESA-----ANSLEWLADLSEK--HMDHLNPLPVET-----LKSDDDKGKEREQADA 149
Db 292 TPKTSDHMETPSGDQWITIDYKQKFVSEYFHREPVVPEMGGFLYKSDGRRSMK----- 345
QY 150 KYKAFQOLYGDSESSILYAPVQLPLVGEVT---FFDQSAERKGEISQLKSMLTITV 205
Db 346 --KHYPVL---RPSGLYYPAPKSKKPTTKDITCLMNLHNSNQVYTGIGMEKKYKSPTPWCI 399
QY 206 AQERFATQFMENAKRCVQOLDRLSA---LVSTKCHSLGSOSTNFGFAKSLITRVENALV 262
Db 400 STKLTALQMRKRSOFIKYICAEDEMTFRKWLVALRIAKNGAE-----LLENYERA-C 449
QY 263 HLSGIKLAPKAFAKTVQEAESSVSEGLPSHMDTKHIERIPMAEQOQTVSOHLHAGN 322
Db 450 QIRRETLGPAS-----SMSAASSSTAISEVPHSL--SHQRTPSVASSIQ-LSSHM----- 497
QY 323 LSELGNLNNNRDLAPHLRLREVSDYFRQSEPHSPISFLEKAIKMGYLSLPELLREMMSE 382
Db 498 -----MNPETHPLSVVRNQ-----SPASFVNSCOQ---SHPSRTSAKLEI 536
QY 383 QNGDALSTIFNAGLNLDOVLLPEVSTPVVGIES-----PQTP-----QAKPSYSD 429
Db 537 QYDEQPTGTIKRAPLDVLRKVRASSTSSPTIQEESDSDDEFPAPPVAVSMRMPVPVT 596
QY 430 PRSVEEHVSQTSF-----VDTOSKODOKPOSSATSAL 461
Db 597 PKPCTPLTSKKAPPPPKRSDTTKLSASAPMAKANDL 634

```

Search completed: March 2, 2003, 04:19:12  
 Job time : 39 secs

GenCore version 5.1.4.p5 4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2003, 21:25:27 ; Search time 72 seconds  
(without alignments)  
1324.997 Million cell updates/sec

Title: US-09-915-706a-2

Perfect score: 2343

Sequence: 1 MPLSKHQIQGLSKPLSDSI.....DTQSKQDKPKQSATSALSM 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL\_21:\*  
2: sp-archaea:\*  
3: sp-bacteria:\*  
4: sp-fungi:\*  
5: sp-human:\*  
6: sp-invertebrate:\*  
7: sp-mhc:\*  
8: sp-organelle:\*  
9: sp-phage:\*  
10: sp-plant:\*  
11: sp-rodent:\*  
12: sp-virus:\*  
13: sp-vertebrate:\*  
14: sp-unclassified:\*  
15: sp-virus:\*  
16: sp-bacteriaph:\*  
17: sp-archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	252	10.8	358	2 Q93ED3	Q93ED3 rhizobium 1
2	241.5	10.3	351	16 Q807V6	Q807V6 agrobacteri
3	232.5	9.9	337	16 Q8XRS2	Q8XRS2 ralsstonia s
4	209	8.9	366	16 Q91IB9	Q91IB9 pseudomonas
5	152.5	6.5	1454	4 Q9UPS3	Q9UPS3 homo sapien
6	143.5	6.1	959	5 Q8TJP2	Q8TJP2 dictyosteli
7	143	6.1	884	5 Q02239	Q02239 caenorhabdi
8	141.5	6.0	344	16 Q91750	Q91750 pseudomonas
9	141.5	6.0	1501	5 Q22699	Q22699 caenorhabdi
10	140.5	6.0	3742	8 Q9TKW8	Q9TKW8 nephroselimi
11	136.5	5.8	7201	5 Q9VMT8	Q9VMT8 drosophila
12	134	5.7	1051	3 Q60039	Q60039 emericella
13	129.5	5.5	630	5 Q25613	Q25613 onchocerca
14	129.5	5.5	1133	5 Q21022	Q21022 caenorhabdi
15	128	5.5	1978	4 Q15154	Q15154 homo sapien
16	127.5	5.4	2471	12 Q9YTK3	Q9YTK3 ateline her

17	127	5.4	1935	6 Q9BE39	Q9BE39 bos taurus
18	126.5	5.4	752	4 Q13597	Q13597 homo sapien
19	126.5	5.4	1023	16 Q92B64	Q92B64 listeria in
20	126	5.4	613	2 Q9XD53	Q9XD53 moraxella c
21	125.5	5.4	1992	13 Q04834	Q04834 xenopus lae
22	125	5.3	1194	3 Q42649	Q42649 schizosacch
23	125	5.3	1935	6 Q9GKR1	Q9GKR1 sus scrofa
24	124.5	5.3	550	3 Q59725	Q59725 schizosacch
25	124.5	5.3	1964	13 Q93522	Q93522 xenopus lae
26	123.5	5.3	1051	2 Q49524	Q49524 mycoplasma
27	123	5.3	611	4 Q14905	Q14905 homo sapien
28	123	5.2	800	10 Q94GM1	Q94GM1 oryza sativ
29	123	5.2	1935	4 Q9H1D5	Q9H1D5 homo sapien
30	123	5.2	2471	13 Q9DDN8	Q9DDN8 xenopus lae
31	122	5.2	1344	2 Q49545	Q49545 mycoplasma
32	122	5.2	1370	5 Q9VD13	Q9VD13 drosophila
33	122	5.2	2735	5 Q76416	Q76416 caenorhabdi
34	121.5	5.2	1179	16 Q97QG7	Q97QG7 streptococc
35	121.5	5.2	1871	5 Q9NCL3	Q9NCL3 drosophila
36	121.5	5.2	3259	4 Q14789	Q14789 homo sapien
37	121	5.2	765	11 Q9D2N7	Q9D2N7 mus musculu
38	121	5.2	1058	10 Q9FLR5	Q9FLR5 arabidopsis
39	121	5.2	2501	5 Q9NCW7	Q9NCW7 rattus norv
40	120.5	5.1	862	11 P70609	P70609 drosophila
41	120.5	5.1	1048	5 Q9W018	Q9W018 drosophila
42	120.5	5.1	1294	5 Q9VXU1	Q9VXU1 drosophila
43	120.5	5.1	1398	5 Q96DD0	Q96DD0 drosophila
44	120.5	5.1	1939	6 Q9TV61	Q9TV61 sus scrofa
45	120	5.1	391	16 Q98IM6	Q98IM6 rhizobium 1

#### ALIGNMENTS

RESULT 1  
ID Q93ED3 PRELIMINARY; PRT; 358 AA.  
AC Q93ED3:  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE IMPA.  
GN IMPA.  
OS Rhizobium leguminosarum (biovar trifolii).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID:386;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Bladergroen M.R., Badelt K., Stronk O.P., Lugtenberg E.J.J.,  
RA Spalnk H.P.;  
RT "An avirulence locus of a symbiotic Rhizobium leguminosarum strain is  
RT involved in temperature-dependent protein secretion."  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF61470; AAL17808.1; -  
SQ SEQUENCE 358 AA; 39039 MW; 9F5C107C58D22395 CRC64;

Query Match 10.8%; Score 252; DB 2; Length 358;  
Best Local Similarity 23.9%; Pred. No. 4.1e-09;  
Matches 96; Conservative 71; Mismatches 170; Indels 64; Gaps 14;  
QY 8 IEQLSKPLSDSICGVYLLKLEKSAFRLRNENFAQTALRLSQNPASDERDAIQEACLN 67  
Db 3 VREITIDPLEONHPCGEVNR-SNTAFREIYRIKARNAARAEKSIIPGETITTAIPA-- 58  
QY 68 KWKILSDSLVEQFSKTRTRDELISMPVAAGFLDPTLESANSLSEWLAD-----LSEK 120  
Db 59 -WHQVSNIGUQLLSKSKDIEVLAWLAQAQLRL-----RGPSGLHDVYVATVSLDK 109  
QY 121 HMDHLNPLYLVEYETLSKDDCKEKEREQADAKYKAFQVLGDSSESSIIYAPYQLPLVG-- 178  
Db 110 HFDALHSI-----GDGD-----VEERFAPAGANGVGEGTLIAIRLTSLIPGCK 155

```
QY 179 --EYTFDFQSAERKGEISOLKSMITTTVAQERFAIOFKMENARCVTQDLRLSALVSTK 236
DB 156 FAQSFSLWDFQISQRPNEFERRERKLEQAAAGVAMSAVLDVLTGCLAAFPRIWEIIDEH 215
QY 237 CHSLGOSTNFGFAKSLITREYNALVHLSGIKLAPKAEAKTVEQEVSESSVSEGLPSHM 296
DB 216 CGDQAPFSSN---TRNVLOEASAIRMLAGIE-AP-TEAPLAQALASQSE-PLPA-S 268
QY 297 DTKHIERIPMAEQAVTSQHLHAGNLSLGNLNMNDLAFHLLREVSDFRQSEPHSP 356
DB 269 ETDSESAIGLRPSAEMT-----RSREAFPELLIAVAFRFRREPHSP 311
QY 357 ISFLLEKAIKRWGYSLEPELLREMMSEONGDALSTIFNAGL 397
DB 312 ISMSIETLVRRGRMDFELLAEILPEQ--QTRNAVLTAAGI 350

RESULT 2
ID 0807V6 PRELIMINARY; PRT; 351 AA.
AC 0807V6;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein ACu4343.
GN Atu4343 OR AGR_L_1042.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxId=176299;
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; Pubmed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitaajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; Pubmed=11743194;
RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Hailing C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009363; AAL45137.1; -.
DR EMBL: AE008251; AAK69097.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 351 AA; 38490 MW; 85460F21E1C63FC7 CRC64;

Query Match 10.3%; Score 241.5; DB 16; Length 351;
Best Local Similarity 23.6%; Pred. No. 2e-08;
Matches 99; Conservative 64; Mismatches 162; Indels 85; Gaps 15;
```

```
DB 49 PQDMLKISSSDSVNGLQITIIYSEKDVLELALAEASRL- RGFHGLREIYELCGDLF 107
QY 119 EKHWDLNLPVLETLKSDDGKREQADAKVAFQLVGDSSESSITLAPVLQPLV- 177
DB 108 YNHWDLSRLSI-----SDND-----EKKFAPFAGLNGISEGT-LVQPLRLASLIP 152
QY 178 ----GEYTFDFQSAERKGEISOLKSMITTTVAQERFAIOFKMENARCVTQDLRLSALV 233
DB 153 GKGKGEHSLWDFQISQRPNEFERRERKLEQAAAGVAMSSHLAAVNTCLSSFPAITAVL 212
QY 234 STKCHSGOSTNFGFAKSLITREYNALVHLSGIKLAPKAEAKTVEQEVSESSVSEGLPSHM 290
DB 213 SERCGQAPFSSN-----IRNTLIEAA-----AAIRTLGRDQDPAP----- 249
QY 291 ELPSPMDTKHIERIPMAEQAVTSQHLHAGNLSLGNLNMNDLAFHLLREVSDFRQ 350
DB 250 -----VEQTPAIA--AGTDESGQSAARTSPASPEGISSRDEAFETLLSVARFR 297
QY 351 SEPHSPISFLLEKAIKRWGYSLEPELLREMMSEONGDALSTIFNAGL 397
DB 298 TEPHSPISLISLETLVRRGRMDFELLAEILPET--QARNNAVLTAAGI 342

RESULT 3
ID 08XRS2 PRELIMINARY; PRT; 337 AA.
AC 08XRS2;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Hypothetical protein RSP0759.
GN RSP0759 OR RS01949.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxId=305;
RN
RP SEQUENCE FROM N.A.
RX STRAIN=GM11000;
RX MEDLINE=216181879; Pubmed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brothier P., Camus J.C., Catorlico L.,
RA Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Signier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL: AL646080; CAD17910.1; -.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 337 AA; 37094 MW; C1EB68258B839897 CRC64;

Query Match 9.9%; Score 232.5; DB 16; Length 337;
Best Local Similarity 24.9%; Pred. No. 7.7e-08;
Matches 99; Conservative 60; Mismatches 127; Indels 111; Gaps 18;
```

QY 247 GGFASLLTTRVNAIVHLSGTLK-----APKAEKTVQEOVAESSVSGGELPSHNDTHI 301  
 Db 209 FRAREALQAV-RAIVEFGGKTETKPSAVEKATTAPOQTQ----- 250  
 QY 302 ERIPASQAOQTVSOHLHAGNLSELGNLNNNRDLAFHLLREVSDFROSEPHSPISFL 361  
 Db 251 --PVAAEAVPDV-----VGPTR--TRQALNQLRVAEFPFQTEPHSPVAYIA 294  
 QY 362 EKATWGLSLPELLREMMSEONGDA-ISTFPAAGL 397  
 Db 295 ARAAKMGDMPLHMLRAVAVKD---DATLSQIEELLGL 328  
 RESULT 4  
 Q911B9  
 ID Q911B9 PRELIMINARY; PRT; 366 AA.  
 AC Q911B9;  
 DT 01-MAR-2001 (TREMblrel, 16, Created)  
 DT 01-MAR-2001 (TREMblrel, 16, Last sequence update)  
 DT 01-OCT-2001 (TREMblrel, 18, Last annotation update)  
 DE Hypothetical protein Pa2360.  
 GN PA2360.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 CC Pseudomonas.  
 OX NCBI\_TaxID=287;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-ATCC 15692 / PAOI;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-T., Errin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Girdy R.L., Goltzer L., Tolentino E., Westbrock-Wadman S., Yuan Y.,  
 RA Braden L.L., Coulter S.N., Folger K.R., Kas A., Labig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an  
 opportunistic pathogen.";  
 RL Nature 406:959-964(2000).  
 DR EMBL; AE004662; AAG05748.1; .  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 366 AA; 40816 MW; E32CFIDB0116536A CRC64;  
 Query Match 8.9%; Score 209; DB 16; Length 366;  
 Best Local Similarity 21.5%; Pred. No. 3.3e-06;  
 Matches 91; Conservative 73; Mismatches 151; Indels 108; Gaps 16;  
 QY 1 MPLSKH-QIEOLSKPLRSDSICGYLYLLEKSAFPLRNEFVAQTLAKLKSQNSPDR 58  
 Db 3 LPLSNALSLLEVLEPIPGQACGSLKRDPIYR-----LKELRREDDSLP 50  
 QY 59 DALQEAELNK--WKILSDSLYEQFSKTRDIELISWFAAQFLDPTTEASANSLEMLAD 116  
 Db 51 TGVWQAEAKRADMAVEQLASELLQGRSKDMLAMLGEA-WLQRGGLGGLRALVILLAE 109  
 QY 117 LSEKIMHNLNPLPVETLKSDDDKCKEKERQADAKVKAFFQVLGDSSESSILYAVYLQPL 176  
 Db 110 LCERYPEEHN-----QAQDDQSMRWVPPIDWLRLRYAEELH-----TRLP 151  
 QY 177 VG-----EVTFFDFQSAERK-----GEISOLKSMLTYYAQBFAILOFKENAKR 221  
 Db 152 MGQGAFAETTLAYMRQROQVAVSSDSKAKALAAQLOQK-----KLDELRAREPLVQ 206  
 QY 222 -----CVTLQ-----DLRLSAVSTKCHSLGSSQTFNGFAKSLITRVEVALVHL 264  
 Db 207 WQRQASLLACQQLRLRLBQWCDRLGLGLARSCQPLREV-----IAQWALLKEFLAMH- 260  
 QY 265 SGCIAPKPAEAKTVQEOVAESSVSGELPSHMDTHIERIPMASQAOQTVSOHLHAGNLS 324  
 Db 261 -----POAPLSEEDPPVAEADASG-----DTGGEESVPASASGPAAG----- 300  
 QY 325 ELGNLNNNRDLAFHLLREVSDFROSEPHSPISFLLEKATWGLSLPELLREMMSEON 384

Db	301	-----1SRDANYQQLLLADIYLTARTPHSPVPLIRAVBMGKKPLSELLAEI	---N	350
QY	385	GDA 387		
Db	351	ADS 353		
RESULT	5			
ID	Q9UPS3	PRELIMINARY;	PRT; 1454 AA.	
AC	Q9UPS3;			
DT	01-MAY-2000 (Tremblrel, 13, Created)			
DT	01-MAY-2000 (Tremblrel, 13, Last sequence update)			
DT	01-MAR-2002 (Tremblrel, 20, Last annotation update)			
DE	KIA1079 protein.			
GN	KIA1079.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RX	MEDLINE=99397452; PubMed=10470851;			
RA	Kikuno R., Nagase T., Ishikawa K., Hirotsawa M., Miyajima N.,			
RA	Tanaka A., Kotani H., Nomura N., Ohara O.			
RT	"Prediction of the coding sequences of unidentified human genes. XIV.			
RT	The complete sequences of 100 new cDNA clones from brain which code			
RL	for large proteins in vitro."			
RL	DNA Res. 6:197-205(1999).			
DR	EMBL; AB029002; BAA83031.1; -.			
DR	HSSP; P12931; IPRK.			
DR	InterPro; IPR000719; Euk_pkinase.			
DR	InterPro; IPR004040; STY_pkinase.			
DR	InterPro; IPR001245; Tyr_pkinase.			
DR	Pfam; PF00069; Pkinase; 1.			
DR	PRINTS; PR00109; TYRKINASE.			
DR	ProDom; PD000001; Euk_Pkinase; 1.			
DR	SMART; SM00221; STYKC; 1.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.			
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.			
DR	ATP-binding; Transferase; Tyrosine-protein kinase.			
DR	SEQUENCE 1454 AA; 159543 MW; 86823F1C45AD559 CRC64;			
Query Match	6.5%;	Score 152.5;	DB 4;	Length 1454;
Best Local Similarity	21.4%;	Pred. No. 0.14;		
Matches 122;	Conservative 75;	Mismatches 189;	Indels 183;	Gaps 25
QY	3	LSKHQIEQLSLPLSD---DSICGVILKLEKSAFRLRNEFNVAQTLAKRLSONPSADERD	59	
Db	370	LKPLRLQ---PYSRMLREVLQFCMLSPK---RAAEVHMLLYLRLQSGRDS--EVD	421	
QY	60	ALQEACLKMKILSDLYEQFSKTTTRDIELISWYAAQF-----LDDTTLESANSL	113	
Db	422	FEQQ-----WNAALKPNTNSRDSSNNAAPILIDHFAARDRGREMEELVTETSSQISFEY	476	
QY	114	LADLSE-KHMD-----HIN-----PVLRYETLKSDDDKGKEREDADKVAFAFL	157	
Db	477	VWEAKKHDFDRSGCHLDEGLSTYSIFYVEVFESSUSDGPGKQ-----	522	
QY	158	VDSSESSILVAPVQLPLVGEVTFPD-----FQSAERKG---EISQLKMLTT	203	
Db	523	-DDSGQDVPPLRP-----GVPPYFDANHLNSVSGDYIQLLEKSGSNLELDYPPALLTT	574	
QY	204	TYAQERFAIQERENAKRCVQDRLSLALVSTKCHSLGSGSTNFGFANSLTRVENALVH	263	
Db	575	-----DMDNPERKGPQLSQTALRSVE---LEESSTDEDFQFSSDTPDKSSILPG	620	
QY	264	LSGIKLAKAEAKTYEQEAESVSEGGELPPIIMDKHIERIPMADEQAQ---TVSQHLH	319	
Db	621	DLHVTSGESEPNNIIFNDKSE---DLPSHQ--KIFDMLNLGVQADFKATLTSSSID	674	

QY	320	-----	AGNLSELGNLNNNNRPLAFLHRESDYFROSE	352
Db	675	NPKRESVITGHEEKPKPKIIPDSEPLCLSDNLHMQDFDPLN	-----VQELSEN	723
QY	353	PHSPISPLLEKALRWGLSLPELLREMMSE-ONGDALSTIFNAAGLNHLD	-----	401
Db	724	-----LFLQEKNLKGLSLSKEHINLQLETLKNAGFTEAMLETSCNSDNLDTLOFAENKP	778	
QY	402	-----QVLL	-----PEVSTPIVGLIESPOTPAKPSVSDPRV	433
Db	779	GMSLLQENWSTKGDDIDVMLTGLTGLSTGLSSQSPSEVOVPPSTETETPRRVPDPSLPIQV	838	
QY	434	EEHVSQTSPYDTSQKQDQKFOSSATSALS	462	
Db	839	E---TOPPLCDVIVPEEDCLHNDISPAVYT	864	
RESULT 6				
Q871P2				
ID	0871P2	PRELIMINARY;	PRT;	959 AA.
AC	0871P2			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	TVIA element modulatory factor.			
OS	Dictyostelium discoideum (Slime mold).			
OC	Eukaryota; Mycetozoa; Dictyostelidida; Dictyostelium.			
OX	NCBI_TaxID=44689;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AY4;			
RA	Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,			
RA	Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,			
RA	Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,			
RT	Sequence and Analysis of Chromosome 2 of Dictyostelium.			
RL	Submitted (MAR-2002) to the EMBL/Genbank/DBD databases.			
DR	EMBL, ACL16330; AAL96709.1;			
DQ	SEQUENCE 959 AA; 110909 MW; 151551A907D72335 CRC64;			

```

Query Match          6.1%; Score 143.5; DB 5; Length 959;
Best Local Similarity 19.7%; Pred. No. 0.31;
Matches 88; Conservative 78; Mismatches 173; Indels 107; Gaps 16;

QY      41 VAOTALRLKLSQNPSSADE--RDA--LOEACLYKKKILSD-----SLVEQFS 81
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      498 VSDSAKCDLRKDQDQDAIKAIQATLDQALDKAKMDIQDLNKHNSIEDRYKKQIQEERS 557
      : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      82 KTTBIEILISMFVAQFLDITTTLESANSLKMLADLSEKNMDHLNVLPEVFLKSDDGK 141
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      558 RIKSEFOILIANNEKEFL-----DKIEOLFIEIENHRTINNNKEQYWK--EEELN 607
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      142 KE-----REQADAKKAFQOLVGDSEESIIYAPVLQDLYAEVYFFPQASERKEISQ 196
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      608 KEIHLQORCDRAETR-----NDOLSSI---PQATRLPQIOIESIQEENFERSQWTWT 658
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      197 LKSMLTVADEERFAIQFKMENAKRCVYQLDRLSALVSTKSLGSOSTNPFKASILTR 256
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      659 LEKQLNNQDREBRLKAENHVNRQOILSLEL-----TLKFLQLESES-----KIDKK 708
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      257 VENALVHLSGITLAKAEKTYEYQEAVSSSEGLSPSMQTKHI--EKIPMAS----- 308
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      709 LKSTTELQOLKSIQSDHLKSIQDESSSSSTQSKIDQHLNLTISLEERLTLSNKKELDL 768
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      309 -----EOAQVSOHLHAGNLSLSEGNINNNNRDLAPHL 340
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      769 EDKROKEKELYKETSNFLKQKQOILLQOQLOOQOQOQENQONQINNNNNNNNNNGSS 828
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      341 LREVSDFYRQSEHSP-----ISFLLEKAIWGYLS-----LP---ELLRPMSE 382
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      829 LSNSTDKINNGSISIPRPOQONKGLFSGIKRTPSNHDLFHHVSGNGLLPASLFLYLSLSQ 888
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY      383 QNGDALSTIFENAGLNHLDDVLLPEV 408
      : : : : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Db 889 KEGEAMSLQAOVQSLNHSRKKLEDEL 914

RESULT 7
ID 002239 PRELIMINARY; PRF: 884 AA.
AC 002239; 002269;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE E01G4.6 protein.
CN E01G4.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RA SENSE FROM N.A.
RP Sins M.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RA SENSE FROM N.A.
RP MEDLINE=94150718; PubMed=7906398;
RX Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Gardner M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gaxton A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Spratt J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RT Nature 368:32-38(1994).
RN [3]
RA SENSE FROM N.A.
RP Sins M.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 283223; CAB05719.1; -.
DR EMBL; 281538; CAB05719.1; JOINED.
DR EMBL; 281538; CAB04387.1; -.
DR EMBL; 283223; CAB04387.1; JOINED.
DR InterPro; IPR001211; PhospholipaseA2.
DR PROSITE; PS00118; PA2_HIS; UNKNOWN.1.
SQ SEQUENCE 884 AA; 99603 MW; 72A1610965C9B30D CRC64;

Query Match 6.1%; Score 143; DB 5; Length 884;
Best Local Similarity 20.7%; Pred. No. 0.3;
Matches 103; Conservative 99; Mismatches 165; Indels 130; Gaps 25.

QY 27 LEKSAFRLRDNFVNAOALRKLKISONSADERDALQEOCLNWK-----KILSDLYEFSK 82
Db 218 LRPDAVITARITRRLATLSTATLSR-----REKFLKYSIROYVAQNLCGOOR 268
QY 83 TTRDIELISWEVAAOFLDITLLESANSLLEWLADISEKHMHDLPVLPYEYLSDDDKGK 142
Db 269 LTRPEYEDLVTESSKSLAVEAPEETTQDETTDASE-----VTTTKAVEATE 317
QY 143 E-RECAADAKVAFAFQVODSESSSLIVAPVLO-----LPLVGEVTFPD-FQ 186
Db 318 EVTEATEATEA---PVATTKESSSEMHVNTIRNMIRSAEKKDLSKYVYLISGKFSLEF 374
QY 187 SAERK-----GETISOLKSMLTITTYAO-----ERFAIOPKMEAKKRCYQL 226
Db 375 LAEOQKLTLLTSKFEKDEKSSKMAKLLDLINALEUSEKSGEIQAMEKKEPKSELVAME 434
QY 227 DRLSLATVSTKCHLSQSTNGFAKSLITR-----VENALVHLSGILKPKAEAKTVEOE 281
Db 435 DKDPAVAVTTIDSLKHKAEKLAFTIYSRVNVEAENM---IEKEVPEKKAEEKVKVED 491
QY 282 ---VAESSVSE---GELPSHMDTKHIERIPMASEQAQTYSOHLHAGNLSLGCNLNM--N 333

```



```
DB 492 VKAAVEKKKKKKPKLP--MKIEKLEK--FVDTK-----SENHBLKVLDD 534
QY 334 RDAFHLRLRVSDYFROSEPHSPISFLLEK---AIRMGYLSLPELREMSSEQ---NGDA 387
DB 535 KERALLVESITKTAETKPKVE-SFRSEETVVAID---DMPALKEESAERKETTGP 589
QY 388 LSTIFNAAGLNHLDOVLLPEVSTPTVIGIESPOTPAKPSVSDRVSVEHVSQTPVDTQ- 446
DB 590 TTT-----EAAVETTEASTPKPEAKPELLS--NLEDVLTITTP-ETET 630
QY 447 ---SKODKPOSSATSA 460
DB 631 IEGSGEREETPTTSAPAA 647

RESULT 8
QY 091750 PRELIMINARY; PRT; 344 AA.
ID 091750;
AC 091750;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE Hypothetical protein PA0082.
GN PA0082.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NCBI_TaxID=287;
RN NCBI_TaxID=287;
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stoyer C.K., Pham X.-Q.T., Ervin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gardner R.L., Goitly L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
DR EMBL; AEO04447; MAG03472.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 344 AA; 36679 MW; 3321BAC4C8DB025D CRC64;

Query Match 6.0%; Score 141.5; DB 16; Length 344;
Best Local Similarity 21.1%; Pred. No. 0.1;
Matches 89; Conservative 70; Mismatches 144; Indels 119; Gaps 18;

QY 15 LSDSDICGVYLYLKEKSAFRLRNEFNVAQTALKKLSQ--NPSADBRALQACLNKKKILS 73
DB 11 VSPDSEPCGDDL-----EYDAALFELERIAOGQPEROMGAVLRAEPPEPRVR 58
QY 74 DSLYEQFSKTTDIELISWFAAFLDITLESANSLEWLADISEKHMDLNPVLPVET 133
DB 59 ALASELFGK-SKDLRVANLLLOSVALD-GLDGLADGILLVRELIGQYMDGVPRL----- 111
QY 134 LKSDDDKGEREQADAKVKAFFOLVGDSESSITLYAVLQL---PLVGEVTF--FDFOQS 187
DB 112 LDADD-----NDPTFRINALTGLVAE-----PLLDLWVAIPLVRSRAGPVLRA 157
QY 188 AEKGEISQLSKMLTTTVAOERFAIOFKMENAKRCYTQDLRLSALVSTKHSIGSOSTNF 247
DB 158 ALNAGLQRPAS---ETLSEQITAGAFADADA-----DALATRALDGADEHAL 204
QY 248 GFAKSLTRVENA-----LVHL-----SGIKLAPKAEAKTYEQEVAES 285
DB 205 AITSGVAERYSAGQGLDGLRLQLRALQVFLYGRPOGAGESLAGAER--VADEOVGA 263
QY 286 SYVEGELPSIMDTKHIERIPMASEQOATVYSQHLHAGNLSELGNLNNKRDLAHFLLREVS 345
DB 264 PVAAVAAPA-----PRAS-----GEIA-----NREDVLRQLDRLL 293
```

```
QY 346 DYFROSEPHSPISFLLEKAIRMGYLSLPELREMSQNDALSTIFNAGLNHLDOVLL 405
DB 294 EYVVRHPPSPVYVLLKRAKTLVTADPAEIVRNLPD-----GISQFETLRG 340
QY 406 PE 407
DB 341 PE 342

RESULT 9
QY 022699 PRELIMINARY; PRT; 1501 AA.
ID 022699;
AC 022699;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical 173.3 kDa protein.
GN T23F2.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OC NCBI_TaxID=6239;
RN NCBI_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None.
RT "Genome sequence of the nematode C. elegans a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [12]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Du Z.;
RT "The sequence of C. elegans cosmid T23F2.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [13]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39649; AAB0384.2; -.
KW Hypothetical protein.
SQ SEQUENCE 1501 AA; 173309 MW; 271052697CE0CF97 CRC64;

Query Match 6.0%; Score 141.5; DB 5; Length 1501;
Best Local Similarity 20.0%; Pred. No. 0.79;
Matches 112; Conservative 89; Mismatches 168; Indels 191; Gaps 26;

QY 9 EQLSKPLSDSDICGVYLYLKEKSAFRLRNEFNVAQTALKKLSQNPASADERDALQACLNK 68
DB 619 QQLSEHESD-----FLSTQOT--LHDLKKEYSAKNNTLL-----VDKFEVEEILLAK 663
QY 69 WKLLSDSLYQFSK---TTRDIELISWFAAFLDITLESANSLEWLADISEKHMDH 124
DB 664 TELV-DALTKQLLENIRKQQRLELS-----LKQSERDQYKKSLSEMTFTAER---- 708
QY 125 LNVLPVETLKSDDDKGEREQADAKV---AFQO-----LVGDS----- 161
DB 709 -VPILEAEILQLSKDK---NETTARKKHQDEYFEDEKLALNDNNIKKERDYLTEHI 763
QY 162 --ESSITLYAVLQLPLVGEVTFPFDOASER--KEGISQLSKMLTTTVAOERFAIOFKMEN 218
DB 764 RANESMI-----ERLKEITSGLKDKLEENOKMQAHLO---KSEL 798
QY 219 AKRCYQOLDRLSALVS-----TCHSISGSOSTNPGFAK----- 251
DB 799 EKKLLSIDHVSQLOSRVNHSDQDVCEQALPIRINKYVGCKPWNKKEITIEGALPDEN 858
QY 252 -----SLLTFRVENALVHL-----SGIKLAPKAEAKTYEQEVAES 286
```

Db 859 EERLRICKAELETTTRQVTVLQOKLVSIIQOQSSQIKRKRIAVEDSNKNTVHTEDESK 918  
Qy 287 VSEGBELPSHMDTKHIERIPMAPEOACTVSOHLNAGNLSELCN-----LNMMNR 334  
Db 919 MKREVEL---KNTLEMERID-SLEAEFVASSIEKSRQKLVNEDNLKOKLDNDMSYSK 974  
Qy 335 DLAF-----HLREVSDFRQSEPHSPISFLEKAIKRWGLSLE-LLREMMSEON-GD 386  
Db 975 EKQWLQWRISNLEKDNSELQKQIOPSS-----EKSL-----SLNKTKLRKTNSEPDGCD 1024  
Qy 387 ALSTIFNAAGLNHLDDVLLPEVSTPTVYGIESPQ-----TQAKPSVSDPR 431  
Db 1025 DMSSTEGGAGSTNSADPMFVSVTAAPVLSTKSPQPSQLADVLNLVRSDEQVLTLEIPEE 1084  
Qy 432 SVEEHVSQTSPPVDQSKODQ 451  
Db 1085 AAKQEEPMNSLEKSTKNKVQ 1104

## RESULT 10

Q9TKW8 PRELIMINARY; PRT; 3742 AA.  
ID O9TKW8  
AC O9TKW8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
DE Hypoetical chloroplast RF2.  
GN YCF2.  
OS Nephroselmis olivacea.  
OC Chloroplast.  
OC Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae;  
OC Chlorodendriales; Chlorodendraceae; Nephroselmis.  
OX NCBI\_TaxID=31312;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99398694; PubMed=10468594;  
RA Turmel M., Otis C., Lemieux C.;  
RT "The complete chloroplast DNA sequence of the green alga Nephroselmis  
RT olivacea: insights into the architecture of ancestral chloroplast  
RT genomes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 96:10248-10253(1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Turmel M., Otis C., Lemieux C.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF137379; AAD54848.1;  
DR InterPro: IPR003593; AAA\_ATPase.  
DR InterPro: IPR003959; AAA\_ATPase-centr.  
DR Pfam: PF00004; AAA\_1.  
DR SMART: SM00382; AAA\_1.  
DR ATP-binding; Chloroplast.  
SQ SEQUENCE 3742 AA; 436427 MW; 1F45143EDBC89275 CRC64;

Query Match 6.0%; Score 140.5; DB 8; Length 3742;

Best Local Similarity 20.7%; Pred. No. 3.3; Indels 167; Gaps 23;  
Matches 101; Conservative 65; Mismatches 154;

Qy 9 EQLSKPLSDSICG---VYLKLEKSARPL-----RNEFNVAQTLRKLSQNPASDEPDAL 61  
Db 283 EELQKQWEYVSYLGPNSYIEL-ITGRFRYGKQKQREQQAIDAKKLAQKQKEEQSQ 341  
Qy 62 QEACLNK-----WKIISLSLEYQSKTTTREDIELISFVAAQFLIDTTLESANSL- 112  
Db 342 KEKLARLDQGNV-----FDKKQTKVELKALMAEWYKEDQEAQAKDLEHKLN 391  
Qy 113 ---WLADLS-----EKHW-----DHLNPLVLP-----VETIKSDDK 140  
Db 392 RQAFATLTMDKVAEDRYEESYLQYPVVGHKIDLSLSPDLPSMTNRNDELNSVSEA 451  
Qy 141 GKREHQADAKVAFQVGVGSESSILYAPVLQPLVGEVTFPFQSAERKGETSOLKSM 200  
Db 452 LKNRRRDAHPIELHRTIGTERLESV--QFML-----EFT--EFAINER-----IRTL 496

Qy 201 LTTTVAQERPAIOF--KMENAKRCVTQDLRLSALVSTKCHLSGSOSTNGEAKSLTRVE 258  
Db 497 YTTSLSSSHADQFLVEQETLLQHLROLDEV-----AKELVDTR 536  
Qy 259 MALVHLSGITLAKRAEKTYEQEVAESVSEGBELPSHMDTKHIERIPMAEQOQVSOHL 318  
Db 537 GSULD--DGIETLEPLLSRREE-----VSRV-MSLEDYERLQNL 573  
Qy 319 HAGNLSELGNMNMNRDLAFHLREVSDFRQSEPHSPISFLEKAIKRWGLSLELPIRE 378  
Db 574 -----RQDSTSP-----HLSKLTISA 591  
Qy 379 MMSEQNDALSTFTPMAGNLHLDQVLLPEVSTPTVYGIESPQTQAKPSVSDPSVEBHS 438  
Db 592 VLNSPSPPLMQGI---PATNPMLLVYRGDIKSPVGLRAGLTG--RPSREAPDSSHNWI 646  
Qy 439 QTS-PVD 444  
Db 647 ITSLPVD 653

## RESULT 11

Q9VMT8 PRELIMINARY; PRT; 7201 AA.  
ID Q9VMT8  
AC Q9VMT8;  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE MSP-300 protein.  
GN MSP-300 OR CG18251 OR CG18252.  
OS Drosophila melanogaster (fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Pphyroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazer J.R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Abril J.F., Abmayri A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Broksstein P., Brotlier P.,  
RA Burlis R.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,  
RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.M., Nelson K.A., Nixon K., Nuskern D.R., Pacleby J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeter F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter R., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weisenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003608; AAF52223.2; .  
DR HSSP: P46939; 1BDH.  
DR Flybase: FBgn0010070; Msp-300.  
DR InterPro: IPR001715; Calponin-like.  
DR InterPro: IPR003445; CytC\_heme\_bind.  
DR InterPro: IPR02017; Spectrin.  
DR Pfam: PF00307; CH; 2.  
DR Pfam: PF00435; Spectrin; 39.  
DR SMART: SM00033; CH; 2.  
DR SMART: SM00150; SPEC; 36.  
DR PROSITE: PS00021; CH; 2.  
DR PROSITE: PS00190; CYTOCHROME\_C; UNKNOWN.1.  
SQ SEQUENCE 7201 AA; 826597 MW; 3084D2637DE94F2 CRC64;

Query Match 5.8%; Score 136.5; DB 5; Length 7201;  
Best Local Similarity 22.0%; Pred. No. 15;  
Matches 132; Conservative 90; Mismatches 178; Indels 201; Gaps 37;

QY 30 SAFRPLRNF-FNVAQTA---LKKLS-----QNPSSAD---ERD---ALQEA---CLNKK 70  
DB 1263 SEFVQDKNDFSMLOTAETELKSLTPLOTQPKNVSDLSKRDLANVQLQASHQLPLH 1322  
QY 71 ILSDSLYEQFSKTRTDIELISWFAAQFLDITLESANSLLEMLADLSKHMNH----- 124  
DB 1323 ALKSELAPLAPDKRPT-LEKEVTEYEMKFEVMEHVKNRGVGLSEYSAK-NNNYTRILA 1380  
QY 125 -----LNPLP---VETLSDDDKGKEREQADAKVKAFFQLYGD-SESSILYAPVQLP- 175  
DB 1381 ELQEWANKAPKINIEALQSEDLTPREYV---VKVQAFKRILDGRMKQDILLADASDELNP 1437  
QY 176 LVGEVTFEFOQSAERKGEISQLKSMLT----- 203  
DB 1438 KEGNIA---EAKRLKGEITKLOEVLASINRVNDHQAVQEDLVNMQOFOAGLIQIKPA 1493  
QY 204 -----TVAOERFALQFMENNAKRCVTOIDRLSALVSTKCHSLGSO 243  
DB 1494 VQOSEKVVNNVSKPLSLEAVAMQMOQFETQ---COEOLDKLHG-TSNISHMLCK 1548  
QY 244 STNFGFAKSLTRVEN--ALVHLSGIKLAPKAEA---KTVEQEA--ESSVSEGE--- 291  
DB 1549 -TN--APDELAMHSRMAVHENAQAQSAKLEKLVANMKSPDADAKLEDVWQGEQOM 1604  
QY 292 --LPSHMDTKHIERI-----PMASDAQTV-----SQHLHAGNLSL- 326  
DB 1605 SRPVALNPHIDKLEKELVTKSFENNEISODQAKLVTLGQNAQDITSLHAFEGAAALKD 1664  
QY 327 -----GNLNNNRDLAFHLREVSDF--ROSEPHSPISFLLEKAIIRNGYLSLPELLRE 378  
DB 1665 RYNQMKGKILQKLEATRGH-INEVSDALISRODFNAKLNF---SNM-----MGOLFN 1713  
QY 379 MMS---EQNGDALST---IFNAGLNHLDO-----VLREVPSTPVGIGSPOTQAKPSV 427  
DB 1714 QTVQVEINPEVEVETSLVHTHLLQEHADKKPSFNAIYDEVQALG---ATPESNML 1769  
QY 428 SDP-----RSVEHVSQTSV-----DTOS-----KODOKPOSSATSL 461  
DB 1770 NDAYTALVNVYONLETNMLQKAALEKWTLLGKMKNDTESHLYLKHQLDKREGPAAEEL 1829  
QY 462 S 462  
DB 1830 S 1830

RESULT 12  
060039 PRELIMINARY: PRT: 1051 AA.  
AC 060039:  
DT 01-AUG-1998 (Tremblrel. 07, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)

DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)  
DE Anucleate primary sterigmata protein APSB.  
GN APSB.  
OS *Emeticella nidulans* (*Aspergillus nidulans*).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eurotiiales; Trichocomaceae; *Emeticella*.  
OX NCBI\_TaxID=5072;  
RN [1]  
RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
RX MEDLINE=99140441; PubMed=10094631.  
RA Suelmann R., Slevers N., Galetzka D., Robertson L., Timberlake W.E.,  
RA Fischer R.;  
RT "Increased nuclear traffic chaos in hyphae of *Aspergillus nidulans*:  
RT molecular characterization of apsb and in vivo observation of nuclear  
RT behaviour";  
RL Mol. Microbiol. 30:831-842(1998).  
CC -I- FUNCTION: INVOLVED IN REGULATION OF NUCLEAR MIGRATION. MAY BE  
CC -I- INVOLVED IN REGULATING NUCLEAR POSITIONING.  
CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -I- SIMILARITY: SOME TO TRICHOHYALIN.  
DR EMBL: AJ003163; CA05918.1; .  
KW Coiled coil.  
FT DOMAIN 10 200 COILED COIL (POTENTIAL).  
FT DOMAIN 239 285 COILED COIL (POTENTIAL).  
FT DOMAIN 325 743 COILED COIL (POTENTIAL).  
FT DOMAIN 787 824 COILED COIL (POTENTIAL).  
FT DOMAIN 857 878 COILED COIL (POTENTIAL).  
FT DOMAIN 950 1004 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1051 AA; 120996 MW; 926201A9E208847A CRC64;

Query Match 5.7%; Score 134; DB 3; Length 1051;  
Best Local Similarity 19.2%; Pred. No. 1.5;  
Matches 86; Conservative 62; Mismatches 134; Indels 166; Gaps 17;

QY 7 QIEQLSKPLSDSICGVYKLEKSAFRLNFEVNAQTALRKLSQNPSSADERDALQ---- 62  
DB 416 KIEECOR-LTED-----LRTODDNLRALQAEWRSASEGIIIRLE---DAQNNLQRYKA 464  
QY 63 -----EACLNKKKILSDSYEQFSKTRTDIELISWFAAQFLDITLESANSLLEMLAD 117  
DB 465 VQOELQECNEEMESLEKSLYEANSKYOR-----LTVQIESSQNEIAFLRE- 509  
QY 118 SEKHMDLNPVLPEVTLKSDDDDKGEREQADAKVKAFFQLYGDSESSILYAPVQLPLY 177  
DB 510 -----EDDGKIK-----IGDLESLTYRMSLO----- 533  
QY 178 GEVTFEFOQSAERKGEISQLKSMLTTTVAOERFALQF---KMNNAKRCVTOIDRLSALV 233  
DB 534 -----SEKDKTKELEGRH---AEERYORBVGSKERQEVORIMNELNREVSA 578  
QY 234 STKCH---SLGSQSTNFGFASLSTRVENALVHLSG-----IKLAPAEAK 276  
DB 579 KECCRLKKNLSAQELETMTWKERLDELNNKIRETLGDLTGRSSLIANIMKLOKLEST 638  
QY 277 TVQOEVAESSVSEGE-----LPSH-MDTKHIERIPMASQEA-----QTVSOH 317  
DB 639 ALELESTRSLDEKETLLNRRDALLSHGLESKLSLELLEREGQARADKQSTEQALKSH 698  
QY 318 LHA-----GNLSLGNLNNNRDLAFHLREVSDFRQSEPHSPISFLLEKAIIRMGV 369  
DB 699 HQASRTITQNNRSILLELNAARNODRK-----RF 726  
QY 370 LSLPELLREMSRONGDALSTIFNAGL 397  
DB 727 TSLDQOQFREQLNERNMSMLLTIMKRLSGM 754

RESULT 13  
025613 PRELIMINARY: PRT: 630 AA.  
AC 025613:  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE Myosin-like antigen (Fragment).  
OS Onchocerca volvulus.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
OC Onchocercidae; Onchocerca.  
OX NCBI\_TaxID=6282;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90301142; Pubmed=2194123;  
RA Ngozi E., Erondu N.E., Donelson J.E.;  
RT "Characterization of a myosin-like antigen from Onchocerca volvulus.";  
RL Mol. Biochem. Parasitol. 40:213-224(1989).  
DR EMBL: M30398; AAA29413.1; -.  
FT NON\_TER  
SQ SEQUENCE 630 AA; 73919 MW; CF093116A40D889F CRC64;  
  
Query Match 5.5%; Score 129.5; DB 5; Length 630;  
Best Local Similarity 22.7%; Pred. No. 1.5; Indels 81; Gaps 15;  
Matches 84; Conservative 64; Mismatches 141;  
  
QY 35 LRNEFVAQTALRKLSQNSADERDALQACLNKMKILSDSLYEQFSKTRDIELISMFV 94  
DB 239 LRSLNNAERMAQL-QN-----RDSILERNNDMKESDALNNEIDLR-LRD-ELLSVRR 290  
QY 95 AAGFL--DTTSAANSLEMLADISEKHMDLNPVLPVETLKSDDCKEREQADAKV 151  
DB 291 DAKEIRRYNTDLOTARNEIKILSTNNEMSKQLN-----AARDKI 331  
QY 152 KAPFOLVGDSESSILVAPVQLPLVGEVTFPFQSAERKGEISQLKSMLTITVAQERFA 211  
DB 332 NSLAKVTDOONKI-----RDLTGEVHHLBEGELKADAGVANLSELDIT---RER 379  
QY 212 IOFEMNAKRCVTOLDRLSALVSTKCHSLGSGSTNFGFAKSLLRVENALVHLSGRLAP 271  
DB 380 IHLLEGONASLDQTELNKIK-----GDIDSLFG-----ENDM-LTKAESN 418  
QY 272 KAEKTYEQEVAESSVEGELPSHMDTKHIERIPMASEOQOTYSOHLHAGLSGLQNLN 331  
DB 419 EAELDRLKOKLQRSIENAKKYSDALDKLREYDRL-----QNLRYREKIKQAEMLTQ 469  
QY 332 MNRDLAHL-----LREVSDFRQSE-PSPISTFLEKAIKRMGYLSLPELLREPMSEON 384  
DB 470 AVQDLSESLNOSRRELDAIDDKLASEGDRNALNSEYKQLQHEVQPMREQLRK--TDEY 527  
QY 385 GDALSTIFNA 394  
DB 528 QAALSDLVNA 537  
  
RESULT 14  
Q21022 PRELIMINARY; PRT; 1133 AA.  
ID Q21022; Q21071;  
AC Q21022; Q21071;  
DT 01-NOV-1996 (TREMblrel. 10, Created)  
DT 01-MAY-1999 (TREMblrel. 10, last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)  
DE F59A2.6 protein.  
GN F59A2.6.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RL lightning J.;  
RU Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Z34801; CAA84332.1; -.  
DR EMBL: Z66514; CAA84332.1; JOINED.  
DR EMBL: Z66514; CAA91344.1; -.  
DR EMBL: Z34801; CAA91344.1; JOINED.  
DR InterPro: IPR000237; GRIP\_domain.  
DR Pfam: PF01465; GRIP; 1.  
SQ SEQUENCE 1133 AA; 129239 MW; 92015E790A9DE01F CRC64;

Query Match 5.5%; Score 129.5; DB 5; Length 1133;  
Best Local Similarity 21.3%; Pred. No. 3.4;  
Matches 110; Conservative 77; Mismatches 184; Indels 145; Gaps 23;  
  
QY 9 EQLSKPLSDSDISGIVYLKLEKSAFRLPNEFNVAQTALRKLSQNSADE-----RDALQEA 64  
DB 116 EQLL-----AMEKLSNDSQNTLDEVTYKLEQ--SEEEVLAARGAIOE- 155  
QY 65 CLKMKWILSDSLYEQFSKTR---RDIELISFVAQAQFLDPTTESAANSLEMLADISEKH 121  
DB 156 -----LTKLESEKETSTAKTELEAVS-----KKLDSSTLSKESDME-- 196  
QY 122 WDHLNPVLPVETLKSDDCKEREQADAKVKAFFQVGDSESSILYAPVQLPLVGEVTF 181  
DB 197 -----AMKIQLINCEKQKQAEVALLKQKLEVEKKNMSDVE-----YQKQLLEST 241  
QY 182 FFDF-----QSAERKGEISQK-----SMLTTVAQERFAIQ---FKMENA 219  
DB 242 TSEKQHAEEAEIYKQLEEAQSSIEMLKDAENENRLKTALESDESSAISLETKQMEAA 301  
QY 220 KRCV-----TQDLRLSALVSTKCHSLGSG-----STNFGFAKSLLRVENAL 261  
DB 302 KKELEASERKSELREQMDRLQ-----KVHNAQEDIQKLTQWELPMKIAKSTEDKL 356  
QY 262 V-HLSGIKLAPRAEAKTVEQEV-----AESVSSEGLPSHMDTKHIERIPMASEOQOT 313  
DB 357 ARQQLAGELNNAKEDLKVVEEKHTGIGRAQALDADEKVEVYKQLQERASQALSSQOE 416  
QY 314 VSOHLHAGNISEL-GNINMNR-----DLAHLREVSDFRQSEPHSIFSLLEKAIK 366  
DB 417 LASSQKADKIQLEKELONAQKRSSELEFANMVSRLTATLNS--NSSTELLKOK-- 471  
QY 367 WGYLSLPELLREPMSEONCD-ALSTIFNAGLHMLDQVLLPEVSTPVGIESPQTPQAKP 425  
DB 472 -----LETLEKELQARQOTKALTEELINVLTS-----LAKEQQTQIQNLQTIQYQ 520  
QY 426 SVSDPSRSEEHVSQTSFVDTQSKODQKPOSSATSAL 461  
DB 521 EVEKEKVE-----LVKVLQQAQSSSSASEAL 549  
  
RESULT 15  
Q015154 PRELIMINARY; PRT; 1978 AA.  
ID Q015154;  
AC Q015154;  
DT 01-JAN-1998 (TREMblrel. 05, Created)  
DT 01-JAN-1998 (TREMblrel. 05, last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, last annotation update)  
DE T1P230.  
GN T1P230.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97404344; Pubmed=9256431;  
RA Chang K.H., Chen Y., Chen T.T., Chou W.H., Chen P.L., Ma Y.T.,  
RA Yang-Feng T.L., Deng L., Tsai M.J., O'Malley B.W., Lee W.H.;  
RT "A thyroid hormone receptor coactivator negatively regulated by the  
RT retinoblastoma protein.";  
RL Proc. Natl. Acad. Sci. U.S.A. 94:9040-9045(1997).  
DR EMBL: AF007217; AAD09135.1; -.  
SQ SEQUENCE 1978 AA; 227514 MW; AACBAEB45C75FD7E CRC64;  
  
Query Match 5.5%; Score 128; DB 4; Length 1978;  
Best Local Similarity 21.0%; Pred. No. 9.4;  
Matches 126; Conservative 77; Mismatches 191; Indels 206; Gaps 26;  
  
QY 4 SKQIQEQLSPSLSDSDISGIVYKL-----EKS-----AFRPLRNEFNVAQ 43  
DB 397 AENIMRLSLNODNSIAEDNKLKRMTEVLEKESKILSOEKRELOMSLKLINNEYEVIK 456

```

OY 44 -TALRKLS-----ONPSADERDALQEA 64
DB 457 STATROISLSELDHLRLNLEAKEQELNOSISKEFTLIAIEELDRONQATKHMILIKA 516
OY 65 CLNKKILSDSLYEQFSKTTRIEL-----ISWFAAQFLIDTTLSEANSL-- 111
DB 517 QLSKQONEGDSIISKLODINDEKRRVHOLEDDKMDITKELDVQFVLLIQSEVALNDLHL 576
OY 112 --EWLADLSKHMHDHNPVLPVETLKSDDDKGERQADAKKAFOLVGDSESSITLYA 169
DB 577 TKOKLEDKVENLVQDLN-----KSQESNVSIOKENLELKEHIRO-----NEEELSRIRN 625
OY 170 PVLQ-LPLVGEVTFPFDOSEAKERGEISOLKSMLT-----TTVAOERFAIQFKMENAK---- 220
DB 626 ELMQSLNODSNSNFKDTLKEREAVERNLKONLSELEQLNENLKVAFDVKMENEKLYLA 685
OY 221 -----RCYTQDLRLS---ALVSTKCHSLGOSSTNFGFAKSL----TRVENALVH 263
DB 686 CEDVRHOLEECLAGNNQLSLEKNTIVETLMEKEGTEAELCMAKKRLLEANKYEKTEIE 745
OY 264 LSG-----IKLAPKAAKTVEQOVAESSVSEGLPSHMDTKHIERIPMA 307
DB 746 LSNARNLNTSAQLDEHEHLIKLQOKR---DMEIAELKRN---IEQMDTDHKEKTDVL 796
OY 308 SEOAQTVSOHLHAGNLSELGNLNNMRDLAFHLLREVSDFYROSEPHSPISFLEK---A 364
DB 797 SSSLEQOKO-----LTQLIN---KKEIFIEKLKERSKLOEE-----LDKYSOA 837
OY 365 IRNGYLSLPELRE-----MMSEONGDALSTIFNAGLNHLDOVL--LPEVSTPT 412
DB 838 LRKN-----EILROTIEEKDRSLGSMKEEN-----NHLQEELERLREOSRT 879
OY 413 VGIESPQT---PQAKPSVSDPRSVVEHV-----SQTSPVDTQSKQOKPQ 454
DB 880 APVADPRTLDSTVELASEVSQNLNTIKHELEELIKHHQKIIEDONOSKMOLLQSLQEOKE 939

```

Search completed: March 2, 2003, 04:11:38  
Job time : 85 secs

